EVERNINOMICIN BIOSYNTHETIC GENES

FIELD OF THE INVENTION

This invention is directed to nucleic acid molecules which encode proteins that direct the synthesis of the orthosomycin everninomicin. The present invention also is directed to use of DNA to produce compounds exhibiting antibiotic activity based on the everninomycin structure.

BACKGROUND OF THE INVENTION

Everninomicin Biosynthesis

Everninomicin is an oligosaccharide antibiotic belonging to the orthosomycin group of antibiotics produced by *Micromonospora carbonacea var. africana* (ATCC 39149, SCC 1413) and is useful as a human medicine. Everninomicin chemically consists of several glycosyl residues attached to modified orsellinic acid. Everninomicin's antibiotic activity is believed to be due to its inhibition of protein synthesis by a mechanism that involves binding of the antibiotic to a ribosome (McNicholas *et al.*, Abstract C-846, ICAAC, San Francisco, CA, 1999). Everninomicin is structurally similar to the antibiotic avilamycin produced by *Streptomyces viridochromogenes* Tu57.

The biosynthesis and enzymatic steps necessary for synthesis of homologs of the chemical moieties contained in the everninomicin structure have been studied in other systems. These include synthesis of orsellinic acid (Type I polyketide), glycosyl group synthesis (deoxysugars), and glycosyltransferase responsible for covalent attachment of glycosyl groups. Orsellinic acid biosynthesis in *Penicillium patulum* and *Streptomyces viridochromogenes* Tu57 has been investigated (Beck *et al.*, European Journal of Biochemistry, 1990, 192:487-498; and Gaisser *et al.*, Journal of Bacteriology, 1997, 179:6271-6278). Glycosyl biosynthesis has been reviewed (Hung-wen *et al.*, Annual Review of Microbiology, 1994, 48:223-56; Williams *et al.*, "The Carbohydrates: Chemistry and Biology" Vol. 1B, 1980, 761-798; and Johnson *et al.*, Current Opinion Chem. Biol., 1998, 5:642-9), and been studied in the erythromycin biosynthetic cluster (Summers *et al.*, Microbiology, 1997, 143:3251-3262). Glycosyltransferases have been studied in a number of systems (Olano *et al.*, Molecular Gen. Genetics, 1998, 3:299-308; Fernandez *et al.*, Journal of Bacteriology, 1998, 18:4929-4937; and Wilson *et al.*, Gene, 1998, 214:95-100).

Polyketides are synthesized via a common mechanistic scheme thought to be related to fatty acid synthesis. The cyclic lactone framework is prepared by a series of condensations involving small carboxylic acid residues (acyl groups). Modifications of the structure, such as ketoreduction, dehydration and enolylreduction, also occur during the processing. The synthesis is driven by a set of large multi-functional polypeptides, referred to as polyketide syntheses.

PCT Publication No. WO 93/13663 describes the organization of the gene encoding the polyketide synthase of *Saccharapolyspora erythraea*. The gene is organized in modules, with each module effecting one condensation step. The precise sequence of chain growth and the processing of the growing chain is determined by the genetic information in each module. This PCT publication describes an approach for synthesizing novel polyketide structures by manipulating in several ways the DNA governing the biosynthesis of the cyclic lactone framework. In order to adapt this methodology to other polyketides, however, the DNA molecules directing the biosynthetic processing must first be isolated.

Combinatorial biosynthesis with bacterial deoxy-sugar biosynthetic genes has been demonstrated (Madduri *et al.*, 1998, Nature Biotechnology, 16:69-74) with the antitumor drug epirubicin (4'-epidoxorubicin) produced by *Streptomyces peucetius*. The heterologous sugar biosynthetic genes *avrE* from *Streptomyces avermitilis* and *eryBIV* from *Saccharopolyspora* were introduced into an *S. peucetius dnmV* mutant blocked in the biosynthesis of dausosamine, the deoxysugar component of epirubicin. Product yields were enhanced with *avrE* complementation demonstrating heterologous expression of sugar biosynthetic genes in combinatorial biosynthesis. Glucosylation of the glycopeptide antibiotic vancomycin (Solenberg *et al.*, Chem Biol, 1997, 4:195-202) demonstrated that the heterologous glycosyltransferases *gtfB* and *gtfE* from *Amycolatopsis orientalis* expressed in *E. coli* produced glycosyltransferase capable of adding glucose or xylose to the vancomycin heptapeptide. Additionally, expression of *gtfE* from *Amycolatopsis orientalis* in *Streptomyces toyocaensis* resulted in glucosylation of A47934, producing a novel antibiotic. Thus, cloned glycosyltransferases can be used to produce novel hybrid antibiotics by glycosylation. In order to adapt this methodology to other glycosyl synthetic genes or glycosyltransferases, however, the DNA molecules directing the biosynthetic processing must first be isolated.

Orsellinic acid is synthesized by AviM, a Type I polyketide synthetase in *Streptomyces viridochromogenes* Tu57. An acytyl-CoA is used as the "starter" unit and three manonyl-CoAs are used as "extender" units for the synthesis of orsellinic acid. AviM has been

shown to synthesize orsellinic acid by introduction of *aviM* into *S. lividans* TK24 (Gaisser *et al.*, Journal of Bacteriology, 1997, 179:6271-6278). AviM has homology to the *Penicillium patulum* Type I polyketide synthase for 6-methylsalicylic acid (MSAS). The *M. carbonacea* EvrJ protein has homology to both AviM and MSAS and contains polyketide synthetic active site motifs resembling acyl carrier proteins, β-ketoacyl:ACP synthetases, and acetyl-CoA/Malonyl-CoA:ACP acetyltransferases. Thus EvrJ contains motifs necessary for the condensation of malonyl extender units with the starter acetyl-CoA unit.

The *M. carbonacea* EvrI protein has homology to DpsC from from *S. peucetius* ATCC 29050. Purified DpsC has been shown to use propionyl-CoA as substrate and to be acylated by propionyl-CoA at the Ser-118 residue (Bao *et al.*, J. Bacteriol, 199, 181:4690-5). This has led to the proposal that DpsC is responsible for the choice of proponyl-CoA as the starter acyl unit in the biosynthesis of daunorhubicin by acting as an β-ketoacyl:acyl carrier protein (ACP) synthetase three (KSIII), and catalyzes the first condensation of the propionate-starter unit with malonyl-ACP. Thus EvrI may be responsible for specifying the choice of acetyl-CoA as the starter acyl group in orsellinic acid biosynthesis and condensation with the first malonyl extender unit. EvrI contains a possible Cys-127 acylation site to form the EvrI-Cys-S-acetyl moiety. This active Cys is similar to the active Cys found in the *Streptomyces glaucescens* FabH (KSIII) enzyme.

The success in cloning and manipulating biosynthetic pathways for the products mentioned above demonstrates a need in the art to isolate and harness the biosynthetic pathway for everninomicin. Moreover, there is a need to employ everninomicin biosynthesis in the development of novel molecules by combinatorial biosynthesis.

Genetic Manipulation of Actinomycetes

The ability to insert genes into the actinomycete chromosome is important to avoid plasmid inhibition of secondary metbolite production and to allow the construction of recombinants that do not require antibiotic selection to maintain cloned genes. Vectors have been developed for use in actinomycetes that contain att/int functions for site-specific integration of plasmid DNA. The two systems available make use of the att/int functions of bacteriophage phiC31 (U.S. Patent No. 5,190,870) and plasmid pSAM2 (U.S. Patent No. 5,741, 675). However, there is a need for additional vectors with att/int functions for site-specific integration in *M. carbonacea*.

The present invention addresses these and other needs in the art.

SUMMARY OF THE INVENTION

The present invention advantageously provides the DNA sequence for the gene cluster responsible for encoding everninomicin biosynthetic genes, which provide the machinery for producing everninomicin. As a result, the present invention provides the information needed to synthesize novel everninomicin-related compounds based on everninomicin, arising from modifications of this DNA sequence designed to change glycosyl and modified orsellinic acid groups contained in everninomicin.

Thus, in one embodiment, the invention provides a nucleic acid comprising an everninomicin biosynthetic pathway gene product from a *Micromonospora carbonacea*, *e.g.*, encoding a protein as set forth in Tables 1a and 1b, and in a specific aspect having a coding region (CDR) as set forth in Tables 1a and 1b.

The invention further provides expression vectors, host cells, and related methods of expression of protein gene products, comprising the isolated nucleic acids of the invention.

In addition, isolated polypeptides corresponding to an everninomicin biosynthetic pathway gene product are provided. Specific amino acid sequences of the polypeptides are set forth in Figure 11 (SEQ ID NOS:2-88) and Figure 12 (SEQ ID NOS:96-106).

Furthermore, the invention provides modified *M. carbonacea*, in which an everninomicin biosynthetic pathway gene is knocked-out, or, alternatively, over-expressed (or both). Similarly, the invention provides for metabolic engineering of new everninomicin analogs.

A particular advantage of this invention is the discovery of various everninomicin resistance genes, which can be used as selection markers. Thus, the invention provides a vector comprising an *M. carbonacea* everninomicin biosynthetic pathway resistance gene, and related methods of selection of transfected or transformed host cells.

In a related but distinct aspect, the inventors have discovered a *Micromonospora* site-specific integrase. The gene for the integrase can be incorporated in a vector for integration into any actinomycete, and, particularly *Monospora*. Thus, the invention further provides a method for introducing a heterologous gene into an actinomycete chromosome using this particular vector.

These and other aspects of the invention are better understood by reference to the following Detailed Description and Examples.

DESCRIPTION OF THE DRAWINGS

Figure 1. The structure of everninomicin.

Figure 2A-C. **(A)** Map of cosmid clones and subclones that span the whole region of the everninomicin biosynthetic locus and surrounding genomic DNA. Heavy cross-hatching indicates sequenced regions; light cross-hatching indicates regions for which a cosmid restriction map was obtained. **(B)** Restriction map of cosmid pSPRX272. **(C)** Restriction map of cosmid pSPRX256. In (B) and (C), cross-hatched regions have been sequenced and cloned fragments are indicated by clone designations beneath the fragment.

Figures 3A-D. Map of the everninomicin biosynthetic region of *Micromonospora carbonacea var. africana* DNA. Distances in bf are shown relative to the beginning of the DNA region. Open reading frames (ORF) are indicated by block arrows. The restriction sites for *Bam*Hl, *Bgl*II, *Eco*RI, *Kpn*I, *Pst*I and *Xho*I restriction enzymes are indicated.

Figures 4A-B. Proposed biosynthetic pathway for orsellinic acid synthesis by *evrJ* and malonylCo-**A** synthesis by *evbD*. **(A)** Orsellinic acid biosynthesis. **(B)** Malonyl-CoA biosynthesis.

Figure 5A-B. Biosynthetic pathway for D-6-deoxysugar and L-6-deoxysugar biosynthesis by *evrV, evrW,* and *evrX*.

Figure 6. Map of pSPRH830B *E. coli-Micromonospera* shuttle vector.

Figure 7A-B. **(A)** Map of pSPRH840 integrating vector. **(B)** Sequence of integrase gene (SEQ ID NO:89) and deduce amino acid (SEQ ID NO:90).

Figure 8. Map of pSPRH826b insertion plasmid.

Figure 9A-B. Analysis of *M. carbonacea* and *M. halophytica* pSPRH840 insertion site att-B/attP region. **(A)** Alignment of pMLP1 attP region with religation clone edge sequences. **(B)** pMLP1 attP.

Figure 10. Schematic of specific resistance gene-containing fragments for cloning in the pSPRH830 vector.

Figure 11A-AB. Everninomicin biosynthetic pathway locus sequence (SEQ ID NO:1) with deduced amino acid sequences (SEQ ID NOS:2-88).

Figure 12A-G. Everninomicin biosynthetic pathway locus sequence (SEQ ID NO:95) with deduced amino acid sequences (SEQ ID NOS:96-106).

DETAILED DESCRIPTION

Micromonospora carbonacea var. africana produces several antibiotics, including everninomicin, thiostrepton, chloramphenicol and lasilosid. As noted above, the present invention advantageously provides the DNA sequence for the gene cluster responsible for encoding everninomicin biosynthetic genes, which provide the machinery for producing everninomicin. As a result, the present invention provides the information needed to synthesize novel everninomicin-related compounds based on everninomicin, arising from modifications of this DNA sequence designed to change glycosyl and modified orsellinic acid groups contained in everninomicin.

The invention also advantageously provides an *M. carbonacea*-specific integrase gene and integration sites (*see*, Figures 7B, 9A, and 9B). Use of the pMLP1 att/int site specific integration function allows for increasing a given gene dosage and for adding heterologous genes that lead to the formation of new products, such as hybrid antibiotics. This procedure has many advantages over methods involving autonomously replicating plasmids. In particular, a plasmid containing pMLP1 att/int functions would integrate as a single copy per chromosome. Plasmids comprising the site-specific integrating function would introduce the gene of choice into the chromosome of actinomycetes. Vectors lacking actinomycete origins of replication can only exist in their integrated form in actinomycetes. Integrated vectors are extremely stable which allows the gene copies to be maintained without antibiotic selective pressure. The site-specific nature of the integration allows analysis of the integrants.

"Everninomicin" refers to a lipophilic oligosaccharide antibiotic of the orthosomycin family of antibiotics, which contain at least one acidic phenolic hydrogen, and two orthoester linkages associated with the glycosy residues (Figure 1; *see*, PCT Publication No. WO 93/07904). These include for example everninomicin, curamycin, avilamycin and flambamycins (Ganguly *et al.*, J.C.S. Chemical Communication, 1976, pp. 609-611; "Kirk-Othmer, Encyclopedia of Chemical Technology", Vol 2, 1978, Third Edition, John Wiley and Sons, pp. 205-209; Ollis, *et al.*, Tetrahedron, 1979, 35:105-127). These lipophilic oligosaccharide antibiotics exhibit broad spectrum biological activity against gram positive and some gram negative bacteria in various *in vitro* assays, and *in vivo* activity, for example, in animal models such as murine models of gram positive infection.

An "everninomicin (EV) biosynthetic pathway gene product" from a *Micromonospora carbonacea* refers to any enzyme ("EV biosynthetic enzyme") involved in the biosynthesis of everninomicin. These genes are located in the EV biosynthetic locus on the *M. carbonacea*

chromosome. This locus is depicted in Figures 2A and 3. Since everninomicin is only known to be produced in *M. carbonacea*, for the sake of particularity the EV biosynthetic pathway is associated with this microorganism. However, it should be understood that this term encompasses EV biosynthetic enzymes (and genes encoding such enzymes) isolated from any *M. carbonacea*, and furthermore that these genes may have novel homologues in related actinomycete bacteria that fall within the scope of the claims here. In specific embodiments, these genes are depicted in Figure 11 (SEQ ID NO:1; polypeptides designated as SEQ ID NOS:2-88) and Figure 12 (SEQ ID NO:95; polypeptides designated as SEQ ID NOS:96-106). It is noted that the sequences of Figures 11 and 12 are linked (contiguous) or connected such that they are part of the same cluster, *i.e.*, the sequence in Figure 12 precedes that of Figure 11. Moreover, the present inventors have identified specific categories into which many of the genes from the EV biosynthetic pathway fall, including but by no means limited to, orsellinic acid biosynthetic enzymes, sugar biosynthetic enzymes, glycosyltransferases, tailoring enzymes, regulatory enzymes (serine-threonine kinases), and resistance mechanism enzymes (rRNA methylases and transporter enzymes). These categories are discussed in greater detail, *infra*. The gene products are listed in Tables 1a and 1b.

Table 1a. Gene Products and Putative Enzymatic Functions Involved in Everninomicin Production

Gene Product	CDS[1]	RBS[2]	SEQ ID NO.	Enzymatic Function (Protein ACC No; BLAST Score)	Class
evdA length 417aa	(132...1382)*	(1389...1394)*	2	Similarity to hydroxylase (CAA11782; 6.5e-137)	sugar biosynthetic
evdB length 474aa	(1490...2611)*	(2618...2622)*	3	Hexose aminotransferase, dnrJ homolog (daunorubicin) (P25048; 2.8e-65)	sugar NH2 addition
evdC length 413aa	(2622...3860)*	(3867...3870)*	4	Similar to flavoprotein, oxidase (S39965; 4.4e-92)	sugar biosynthetic
evdD length 390aa	(4143...5312)	(4134...4138)	5	DNTP-hexose glycosyltransferase (AAC01731; 4.6e-49)	Glycosyl transfer
evdE length 308aa	(5309...6235)		6	hexose dehydratase (CAA18814; 8.0e-58)	sugar biosynthetic
evdF length 347aa	(6232...7275)	(6226...6229)	7	dNTP-hexose glycosyltransferase (CAB07092; 3.4e-18)	Glycosyl transfer
evdG length 351aa	(7272...8327)		8	unknown	unknown

evdH length 340aa	(8342...9364)	(8333...8336)	9	dNTP-hexose glycosyltransferase (CAA19930; 0.8)	Glycosyl transfer
evdI length 253aa	(9463...10,224)*	(10,232...10,235)*	10	hydrolase (AAB81835; 6.8e-10)	sugar biosynthetic
evdJ length 250aa	(10,424...11,176)		11	unknown	unknown
evdK length 415aa	(11,208...12,455)		12	hexose dehydratase or empimerase (CAB08849; 3.3e-26)	sugar biosynthetic
evdL length 304aa	(12,108...13,022)*	(13,027...13,030)*	13	dNTP-hexose glycosyltransferase (S37028; 0.010)	Glycosyl transfer
evrA length 317aa	(14,410...15,363)*	(15,369...15,373)*	14	hexose epimerase (CAA12010.1; 1.3e-40)	sugar biosynthetic
evrB length 344aa	(15,380...16,414)*		15	hexose oxidoreductase (ACC01734; 1.3e-65)	sugar biosynthetic
evrC length 484aa	(16,419...17,873)*		16	hexose dehydratase (CAA12009; 2.2e-107)	sugar biosynthetic
evrD length 354aa	(17,870...18,934)*		17	GDP-mannose 4,6-dehydratase (BAA16585; 1.0e-88)	sugar biosynthetic
evrE length 510aa	(19,374...20,906)		18	multidrug efflux transporter (CAB15277; 1.4e-59)	resistance mechanism
evrF length 492aa	(21,064...22,542)	(21,056...22,542)	19	similar to non-heme oxygenate/halogenase (CAA11780; 4.3e-58)	orsellinic acid chlorine addition
evrG length 474aa	(22,748...24,172)	(22,736...22,740)	20	oxidase (Q12737; 5.5e-67)	tailoring
evrH length 348aa	(24,177...25,223)*	(25,230...25,233)*	21	unknown (AAB89073; 3.2e-6)	unknown
evrI length 358aa	(25,550...26,626)		22	acyl starter unit fidelity (daunorubicin homology) (AAA65208; 5.7e-56)	PKS acyl Carbon choice
evrJ length 1264aa	(26,685...30,479)	(26,672...26,676)	23	orsellinic acid synthase 6-methylsalicilic acid synthetase (CAA72713; 0.0e)	polyketide synthetase
evrK length 439aa	(30,557...31,876)*	(31,885...31,888)*	24	Na/H antiporter (BAA16991; 2.1e-14)	unknown
evrL length 313aa	(31,941...32,882)*		25	similar to gene essential to heme biosynthesis (BAA12681; 0.0012)	unknown
evrM length 412aa	(33,167...34,405)*	(34,414...34,418)*	26	similar to p450 hydroxylase (S18530; 3.8e-70)	tailoring

evrN length 253aa	(34,449...35,210)*	(35,219...35,221)*	27	methyl transferase (CAB10751; 0.00061)	tailoring
evrO length 314aa	(35,294...36,238)*		28	unknown (BAA20094; 0.56)	unknown
evrP length 242aa	(36,235...36,963)*		29	unknown (CAB05421; 0.00020)	unknown
evrQ length 342aa	(36,998...38,026)*		30	similar to oxidoreductase and heat stress protein (P80874; 7.8e-31)	tailoring
evrR length 164aa	(38,072...38,566)*		31	low similarity to hexaheme nitrite reductase regulator (P30866; 0.0034)	regulatory (methyl transferase)
evrS length 423aa	(38,892...40,163)*		32	dNTP-hexose glycosyltransferase (AAD15267; 1.9e-36)	Glycosyl transfer
evrT length 224aa	(40,216...40,890)*	(40,899...40,902)*	33	similar to L-proline hydroxylase (BAA 20094; 5.5e-7)	tailoring
evrU length 229aa	(40,887...41,576)*		34	methyltransferase (CAB02029; 5.6e-6)	tailoring
evrV length 342aa	(41,679...42,707)*	(42,714...42,717)*	35	dTDP-glucose epimerase (AAB84886; 3.5e-36)	L-dTDP-glucose biosynthetic
evrW length 333aa	(42,810...43,799)*	(43,807...43,811)*	36	dTDP-glucose dehydratase (CAA72715; 5.1e-136)	D-dTDP-glucose biosynthetic (GDH)
evrX length 355aa	(43,799...44,866)*		37	dTDP-glucose synthetase (A26984; 1.2e-118)	D-dTDP-glucose biosynthetic
evrY length 248aa	(45,014...45,760)*	(45,767...45,770)*	38	dehalogenase (P24069; 5.8e-8)	drug resistance
evrZ length 250aa	(45,962...46,714)*	(45,952...45,956)*	39	similar to muramidase/lysozyme (P25310; 1.2e-77)	drug resistance
evsA length 692aa	(47,156...49,234)*		40	serine threonine kinase (BAA32455; 2.0e-76)	regulatory
evsB length 363aa	(51,627...52,715)	(51,620...51,622)	41	similar to proteases	unknown
evsC length 222aa	(52,889...53,557)		42	similar to MAF involved in septum formation (BAA18425; 1.3e-21)	unknown
evbA length 217aa	(53,554...54,207)		43	O-methyl transferase (AAC44130; 8.6e-38)	tailoring; possible resistance
evbB length 251aa	(54,362...55,117)*	(55,125...55,128)*	44	membrane pump, homolog mithramicin resistance (AAC443581; 2.9e-24)	resistance mechanism

evbC length 319aa	(55,135...56,094)*	(56,100...56,103)*	45	membrane pump, homolog mithramicin resistance (AAC44357; 1.0e-69)	resistance mechanism
evbC2 length 209aa	(56,184...56,813)*		46	ankrylin like (AAC44356; 0.0041)	resistance
evbD length 582aa	(56,961...58,709)	(56,947...56,951)	47	acyl-CoA carboxylase (CAB07068; 7.3e-201)	malonyl-CoA biosynthesis
evbE length 479aa	(58,873...60,312)		48	IMP dehydrogenase (CAA15452; 4.1e-165)	tailoring
evbF length 185aa	(60,472...61,029)*	(61,038...61,040)*	49	hypothetical protein Rv0653c, mycobacterium (CAB07128; 3.8e-06)	regulator
evbF1 length 90aa	(61,288...61,560)		50	unknown	unknown
evbF2 length 152aa	(61,610...62,069)	(61,597...61,599)	51	ORFI Streptomyces peucetius (CAA06602; 0.024)	regulatory/ resistance
evbG length 557aa	(62,122...63,795)		52	ABC transporter (Q11046; 2.7e-170)	drug resistance
evbH length 645aa	(63,891...65,828)	(63,884...63,887)	53	ABC transporter (Q11047; 5.6e-166)	drug resistance
evbI length 467aa	(66,469...67,872)*	(67,883...67,886)*	54	lipoamide dehydrogenase (CAA17075; 1.6e-140)	tailoring
evbJ length 151aa	(67,979...68,434)		55	hypothetical protein Rv3304 [Mycobacterium tuberculosis] (CAA17076; 7.6e-40)	unknown
evbK length 321aa	(68,529...69,494)		56	protease synthase and sporulation regulator; homology to resistance proteins streptomyces (029729; 7.3-7)	regulatory
evbL length 249aa	(69,610...70,359)*		57	acetyltransferase/ phosphotransferase	tailoring
evbM	(70,365...71,285)*		58	hypothetical protein Rv 1584c [Mycobacterium tuberculosis] (CAB09085; 0.32)	unknown
evbN length 209aa	(71,289...71,918)*	(71,926...71,929)*	59	hypothetical protein SC3A7.08 [S. coelicolor] (CAA20071; 4.0e-40)	unknown
evbO length 231aa	(72,284...72,979)		60	putative lipoprotein [S. coelicolor] (CAA19252; 2.6e-20)	unknown
evbP length 420aa	(72,933...74,195)*		61	peptidase (CAA17077; 6.5e-88)	unknown

evbQ length 527aa	(74,707...76,290)*		62	methylmalonyl-Coa mutate (BAA30410; 1.8e-149)	acyl precursor biosynthesis
evbR length 696aa	(76,622...78,712)		63	protein serine/threonine kinase note eukaryotic type (BAA32455; 1.1e-71)	regulatory
evbS length 576aa	(78,791...80,521)		64	phosphomannomutase (CAA17080; 5.4e-91)	sugar biosynthesis
evbT length 286aa	(82,073...82,933)		65	hypothetical protein SC5C7.22c (CAA20634; 5.7e-28)	10-28
evbU length 202aa	(83,280...83,888)*		66	glucose-6-phosphate 1-dehydrogenase low BLAST homology (S61167; 0.00039)	unknown
evbV length 193aa	(84,080...84,661)*		67	uracil phosphoribosyltransferase (CAA17081; 5.6e-60)	unknown
evbW length 338aa	(84,890...85,906)*		68	deoxyribose-phosphate aldolase (AAA79343; 1.3e-54)	unknown
evbX length 477aa	(85,909...87,342)		69	aldehyde dehydrogenase (AAB84440; 4.2e-103)	tailoring
evbY length 735aa	(87,422...88,159)	(87,407...87,411)	70	aldehyde dehydrogenase (CAA71003; 3.4e-16)	tailoring
evbZ length 137aa	(88,292...88,705)	(88,280...88,282)	71	hypothetical protein (CAB06141; 1.3e-16)	unknown
evcA length 301aa	(88,716...89,621)		72	hypothetical protein, putative integral membrane protein [Streptomyces coelicolor] (CAB06143; 4.5e-28)	unknown
evcB length 416aa	(89,817...91,067)		73	cytochrome D oxidase subunit I (P94364; 3.0e-65)	tailoring
evcC length 335aa	(91,078...92,085)	(91,068...91,072)	74	cytochrome D oxidase subunit II (CAA71118; 1.9e-15)	tailoring
evcD length 561aa	(92,148...93,833)		75	ABC transporter (CAA22219; 2.6e-107)	resistance
evcE length 613aa	(93,830...95,671)		76	ABC transporter (AAC44070; 3.4e-32)	resistance
evcF length 229aa	(95,729...96,418)		77	unknown	unknown
evcG length 111aa	(96,440...96,775)*		78	unknown (AAB84787; 1.9e-8)	unknown
evcH length 303aa	(96,894...97,805)		79	unknown (CAA17083; 9.2e-5)	unknown

evcI search length 691aa	(98,287...100,362)		80	unknown (CAA19992; 6.0e-6)	unknown
evcJ length 197aa	(100,733...101,326)*		81	putative ATP/GTP binding protein (CAA19989; 7.9e-59)	unknown
evcJ2 length 134aa	(101,328...101,732)*		82	unknown (CAA19986; 8.6e-23)	unknown
evcK length 117aa	(101,803...102,156)*		83	unknown (CAA19991; 1.7e-36)	unknown
evcL search length 1145aa	(102,204...105,641)*		84	unknown (CAA19992; 4.6e-99)	unknown
evcM length 88aa	(105,907...105,641)		85	putitive uridine kinase (CAA19591; 1.0e-9)	unknown
evcN length 358aa	(106,513...107,589)		86	unknown (CAA17085; 7.5e-120)	unknown
evrMR length 320aa	(107,653...108,615)	(107,637...107,641)	87	homology to 23S rRNA methylase for mycinamicin resistance (myrA) (BAA03674; 1.4e-79)	resistance
evrMR2 length 193aa	(108,635...109,216)		88	homology to gene linked to myrA	resistance

5

Table 1b. Gene Products and Putative Enzymatic Functions Involved in Everninomicin Production

ORF1 length 291	(189-1064)*	(1069-1073)	96	Transcriptional regulator Biotinylation H70979; 8e-31	unknown
ORF2 length 528	(1184-2767)*		97	Propionyl-CoA carboxylase T42208; 0.0e	unknown
ORF3 length 297	(2863-3753)*		98	unknown	unknown
ORF4 length 167	(3776-4276)*	(4280-4284)	99	ECF sigma factor T36644; 8e-26	regulation
ORF5 length 281	(4526-5368)*		100	Membrane protein CAB94598.1; 5e-50	unknown
ORF6 length 252	(5392-6147)*	(6152-6156)	101	rRNA methyltransferase AAG32067.1; 4e-49	resistance
ORF7 length 363	(6194-7282)*		102	O-methyl transferase PP42712; 4e-59	modification
ORF8 length 284	(7280-8133)	(8141-8145)	103	unknown	unknown

ORF9 length 355	(8254-9318)	(9324-9328)	104	oxidoreductase AAG05128.1; 3e-51	modification
ORF10 length 309	(9575-10,504)	(9568-9571)	105	unknown	unknown
ORF11 Length 334	(10,584-11,585)		106	deoxyhexose ketoreductase T17473; 1e-49	sugar modification

Legend for Tables 1a and 1b

* CDS, RBS complement on full length biosynthetic locus sequence

[1] CDS is then putative coding sequence.

[2] RBS is the putative ribosome binding site.

[3] GenBank protein database (http://www.ncbi.nih.gov/Entrez/protein.html)

Although the term "enzymes" is used to refer to the EV biosynthetic pathway gene products, such gene products may be proteins with non-enzymatic functions. Such proteins are also contemplated as falling within the scope of the present invention.

An "EV biosynthetic pathway bottleneck gene" is a gene encoding a product whose level limits the rate of synthesis of everninomicin. Examples of such gene products include, though are not limited to, evrJ (involved in orsellinic acid biosynthesis); evrV, evrW, and evrX (involved in dTDP-glucose synthesis); evbD (involved in malonyl-CoA-synthesis, which is required for orsellinic acid synthesis); and oxidases responsible for oxidation of the amino group on the terminal sugar to produce everninomicin that contains a nitrososugar group. Other likely bottleneck genes include those encoding glycosyltransferases (evdD, evdF, evdH, evdL, and evrS) and tailoring enzymes, particularly sugar modification enzymes.

A modified *Micromonospora carbonacea* refers to a microorganisms that has been genetically engineered to over-express or suppress expression of an EV biosynthetic pathway gene product (enzyme). Such genetic engineering and manipulation is described in detail, *infra*. Preferably, to increase the level of production of everninomicin, the modified microorganism overexpresses one or more bottleneck genes. To produce an everninomicin analog or homolog, various tailoring enzyme genes (*e.g.*, evdB, a hexose aminotransferase that produces an amino sugar; evrF, a nonheme halogenase that chlorinates the orsinillic acid; or an oxidase gene that produces a nitrososugar by oxidation of an aminosugar) may be knocked out. Other knock-outs may be made of putative key genes, resulting in all likelihood in blockage of everninomicin biosynthesis. These include the orsellinic acid synthase (evrJ), dTDP-glucose synthases (evrV, evrW, and evrX), and glycosyltransferases (evdD, evdF, evdH, evdL, and evrS). A knockout of the glycosyltransferase that adds the terminal glycosyl group is expected to produce an everninomicin analog lacking the terminal glycosyl group.

Such genetic construction can be replicated in a different actinomycete, such as a *Streptomyces*, as described *infra*, by introduction of all or part of the modified everninomicin biosynthetic pathway described here into such a host cell.

A *Micromonospora carbonacea* "everninomicin biosynthetic pathway resistance gene product" is a protein or enzyme that confers resistance to everninomicin (and related compounds) to a host cell. Expression of such a gene on a vector provides an alternative selection mechanism for transformed host cells *in vitro* or *in vivo*, and thus can be used in molecular biological manipulations of cells independently of the EV biosynthetic pathway. For example, such a vector can be used to select for a transfected or transformed host cell by culturing the cell in the presence of an amount of everninomicin that is toxic to the host cell lacking the vector.

A *Micromonospora* site-specific Att/Int functions consist of an integrase protein and AttP site, *e.g.*, as depicted in Figure 7B (SEQ ID NO:89) and in a specific embodiment encoded by a nucleic acid having a sequence as depicted in Figure 7B (SEQ ID NO:90), that permits site-specific integration of a vector into an actinomyce, and particularly a *Micromonospera*, genome.

General Definitions

As used herein, the term "isolated" means that the referenced material is removed from the environment in which it is normally found. Thus, an isolated biological material can be free of cellular components, *i.e.*, components of the cells in which the material is found or produced. In the case of nucleic acid molecules, an isolated nucleic acid includes a PCR product, an isolated mRNA, a cDNA, or a restriction fragment. In another embodiment, an isolated nucleic acid is preferably excised from the chromosome in which it may be found, and more preferably is no longer joined to non-regulatory, non-coding regions, or to other genes, located upstream or downstream of the gene contained by the isolated nucleic acid molecule when found in the chromosome. In yet another embodiment, the isolated nucleic acid lacks one or more introns. Isolated nucleic acid molecules include sequences inserted into plasmids, cosmids, artificial chromosomes, and the like. Thus, in a specific embodiment, a recombinant nucleic acid is an isolated nucleic acid. An isolated protein may be associated with other proteins or nucleic acids, or both, with which it associates in the cell, or with cellular membranes if it is a membrane-associated protein. An isolated organelle, cell, or tissue is removed from the anatomical site in which it is found in an organism. An isolated material may be, but need not be, purified.

The term "purified" as used herein refers to material that has been isolated under conditions that reduce or eliminate the presence of unrelated materials, *i.e.*, contaminants, including native materials from which the material is obtained. For example, a purified protein is preferably substantially free of other proteins or nucleic acids with which it is associated in a cell; a purified nucleic acid molecule is preferably substantially free of proteins or other unrelated nucleic acid molecules with which it can be found within a cell. As used herein, the term "substantially free" is used operationally, in the context of analytical testing of the material. Preferably, purified material substantially free of contaminants is at least 50% pure; more preferably, at least 90% pure, and more preferably still at least 99% pure. Purity can be evaluated by chromatography, gel electrophoresis, immunoassay, composition analysis, biological assay, and other methods known in the art.

Methods for purification are well-known in the art. For example, nucleic acids can be purified by precipitation, chromatography (including preparative solid phase chromatography, oligonucleotide hybridization, and triple helix chromatography), ultracentrifugation, and other means. Polypeptides and proteins can be purified by various methods including, without limitation, preparative disc-gel electrophoresis, isoelectric focusing, HPLC, reversed-phase HPLC, gel filtration, ion exchange and partition chromatography, precipitation and salting-out chromatography, extraction, and countercurrent distribution. For some purposes, it is preferable to produce the polypeptide in a recombinant system in which the protein contains an additional sequence tag that facilitates purification, such as, but not limited to, a polyhistidine sequence, or a sequence that specifically binds to an antibody, such as FLAG and GST. The polypeptide can then be purified from a crude lysate of the host cell by chromatography on an appropriate solid-phase matrix. Alternatively, antibodies produced against the protein or against peptides derived therefrom can be used as purification reagents. Cells can be purified by various techniques, including centrifugation, matrix separation (*e.g.*, nylon wool separation), panning and other immunoselection techniques, depletion (*e.g.*, complement depletion of contaminating cells), and cell sorting (*e.g.*, fluorescence activated cell sorting [FACS]). Other purification methods are possible. A purified material may contain less than about 50%, preferably less than about 75%, and most preferably less than about 90%, of the cellular components with which it was originally associated. The "substantially pure" indicates the highest degree of purity which can be achieved using conventional purification techniques known in the art.

In a specific embodiment, the term "about" or "approximately" means within 20%, preferably within 10%, and more preferably within 5% of a given value or range. Alternatively, especially in biological systems, the term "about" means within about a log (*i.e.*, an order of magnitude) preferably within a factor of two of a given value, depending on how quantitative the measurement.

The use of italics indicates a nucleic acid molecule (*e.g.*, *enrJ* cDNA, gene, etc.); normal text indicates the polypeptide or protein.

"Sequence-conservative variants" of a polynucleotide sequence are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position.

"Function-conservative variants" are those in which a given amino acid residue in a protein or enzyme has been changed without altering the overall conformation and function of the polypeptide, including, but not limited to, replacement of an amino acid with one having similar properties (such as, for example, polarity, hydrogen bonding potential, acidic, basic, hydrophobic, aromatic, and the like). Amino acids with similar properties are well known in the art. For example, arginine, histidine and lysine are hydrophilic-basic amino acids and may be interchangeable. Similarly, isoleucine, a hydrophobic amino acid, may be replaced with leucine, methionine or valine. Such changes are expected to have little or no effect on the apparent molecular weight or isoelectric point of the protein or polypeptide. Amino acids other than those indicated as conserved may differ in a protein or enzyme so that the percent protein or amino acid sequence similarity between any two proteins of similar function may vary and may be, for example, from 70% to 99% as determined according to an alignment scheme such as by the Cluster Method, wherein similarity is based on the MEGALIGN algorithm. A "function-conservative variant" also includes a polypeptide or enzyme which has at least 60 % amino acid identity as determined by BLAST or FASTA algorithms, preferably at least 75%, most preferably at least 85%, and even more preferably at least 90%, and which has the same or substantially similar properties or functions as the native or parent protein or enzyme to which it is compared.

The terms "mutant" and "mutation" mean any detectable change in genetic material, *e.g.* DNA, or any process, mechanism, or result of such a change. This includes gene mutations, in which the structure (*e.g.* DNA sequence) of a gene is altered, any gene or DNA arising from any mutation process, and any expression product (*e.g.* protein or enzyme)

expressed by a modified gene or DNA sequence. The term "variant" may also be used to indicate a modified or altered gene, DNA sequence, enzyme, cell, etc., *i.e.*, any kind of mutant.

As used herein, the term "homologous" in all its grammatical forms and spelling variations refers to the relationship between proteins that possess a "common evolutionary origin," including proteins from superfamilies (*e.g.*, the immunoglobulin superfamily) and homologous proteins from different species (*e.g.*, myosin light chain, etc.) (Reeck *et al.*, Cell 50:667, 1987). Such proteins (and their encoding genes) have sequence homology, as reflected by their sequence similarity, whether in terms of percent similarity or the presence of specific residues or motifs at conserved positions.

Accordingly, the term "sequence similarity" in all its grammatical forms refers to the degree of identity or correspondence between nucleic acid or amino acid sequences of proteins that may or may not share a common evolutionary origin (*see* Reeck *et al.*, *supra*). However, in common usage and in the instant application, the term "homologous," when modified with an adverb such as "highly," may refer to sequence similarity and may or may not relate to a common evolutionary origin.

In a specific embodiment, two DNA sequences are "substantially homologous" or "substantially similar" when the encoded polypeptides are at least 35-40% similar as determined by one of the algorithms disclosed herein, preferably at least about 60%, and most preferably at least about 90 or 95% in a highly conserved domain, or, for alleles, across the entire amino acid sequence. Sequence comparison algorithms include BLAST (BLAST P, BLAST N, BLAST X), FASTA, DNA Strider, the GCG (Genetics Computer Group, Program Manual for the GCG Package, *Version 7*, Madison, Wisconsin) pileup program, etc. using the default parameters provided with these algorithms. An example of such a sequence is an allelic or species variant of the specific everninomicin biosynthetic genes of the invention. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system.

Cloning and Expression of EV Biosynthetic Pathway Genes

The present invention contemplates analysis and isolation, and/or construction, of a gene encoding a functional or mutant EV biosynthetic enzyme, including a full length, or naturally occurring form of an EV biosynthetic enzyme, and any antigenic fragments thereof from any source. It further contemplates expression of functional or mutant EV biosynthetic

enzyme protein for evaluation, diagnosis, or, particularly, biosynthesis of everninomicin or other secondary metabolic products.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual,* Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (herein "Sambrook *et al.*, 1989"); *DNA Cloning: A Practical Approach,* Volumes I and II (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed. 1984); *Nucleic Acid Hybridization* [B.D. Hames & S.J. Higgins eds. (1985)]; *Transcription And Translation* [B.D. Hames & S.J. Higgins, eds. (1984)]; *Animal Cell Culture* [R.I. Freshney, ed. (1986)]; *Immobilized Cells And Enzymes* [IRL Press, (1986)]; B.ÊPerbal, *A Practical Guide To Molecular Cloning* (1984); F.M. Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994).

Molecular Biology-Definitions

"Amplification" of DNA, as used herein, denotes the use of polymerase chain reaction (PCR) to increase the concentration of a particular DNA sequence within a mixture of DNA sequences. For a description of PCR see Saiki *et al.*, Science, 239:487, 1988.

A "nucleic acid molecule" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules"); or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"); or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix; or "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone; or nucleic acids containing modified bases, for example thiouracil, thio-guanine and fluoro-uracil. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear (*e.g.*, restriction fragments) or circular DNA molecules, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

A "polynucleotide" or "nucleotide sequence" is a series of nucleotide bases (also called "nucleotides") in DNA and RNA, and means any chain of two or more nucleotides. A nucleotide sequence typically carries genetic information, including the information used by cellular machinery to make proteins and enzymes. These terms include double or single stranded genomic and cDNA, RNA, any synthetic and genetically manipulated polynucleotide, and both sense and anti-sense polynucleotide (although only sense stands are being represented herein). This includes single- and double-stranded molecules, *i.e.*, DNA-DNA, DNA-RNA and RNA-RNA hybrids.

The polynucleotides herein may be flanked by natural regulatory (expression control) sequences, or may be associated with heterologous sequences, including promoters, internal ribosome entry sites (IRES) and other ribosome binding site sequences, enhancers, response elements, suppressors, signal sequences, polyadenylation sequences, introns, 5'- and 3'- non-coding regions, and the like. The nucleic acids may also be modified by many means known in the art. Furthermore, the polynucleotides herein may also be oligonucleotides modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent molecules, biotin, and the like.

A "coding sequence" or a sequence "encoding" an expression product, such as a RNA, polypeptide, protein, or enzyme, is a minimum nucleotide sequence that, when expressed, results in the production of that RNA, polypeptide, protein, or enzyme, *i.e.*, the nucleotide sequence encodes an amino acid sequence for that polypeptide, protein or enzyme. A coding sequence for a protein may include a start codon (usually ATG, though as shown herein, alternative start codons can be used) and a stop codon.

The term "gene", also called a "structural gene" means a DNA sequence that codes for a particular sequence of amino acids, which comprise all or part of one or more proteins or enzymes, and may include regulatory (non-transcribed) DNA sequences, such as promoter sequences, which determine for example the conditions under which the gene is expressed. The transcribed region of the gene may include untranslated regions, including a 5'-untranslated region (UTR) and 3'-UTR, as well as the coding sequence.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels

detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A coding sequence is "under the control of" or "operably (or operatively) associated with" transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced (if it contains introns) and translated into the protein encoded by the coding sequence.

The terms "express" and "expression" mean allowing or causing the information in a gene or DNA sequence to become manifest, for example producing a protein by activating the cellular functions involved in transcription and translation of a corresponding gene or DNA sequence. A DNA sequence is expressed in or by a cell to form an "expression product" such as mRNA or a protein. The expression product itself, *e.g.* the resulting mRNA or protein, may also be said to be "expressed" by the cell. An expression product can be characterized as intracellular, extracellular or secreted. The term "intracellular" means something that is inside a cell. The term "extracellular" means something that is outside a cell. A substance is "secreted" by a cell if it appears in significant measure outside the cell, from somewhere on or inside the cell.

The term "transfection" means the introduction of a heterologous nucleic acid into a host cell. The term "transformation" means the introduction of a heterologous gene, DNA or RNA sequence to a host cell, so that the host cell will express the introduced gene or sequence to produce a desired product. The introduced gene or sequence may also be called a "cloned" or "heterologous" gene or sequence, and may include regulatory or control sequences, such as start, stop, promoter, signal, secretion, or other sequences used by a cell's genetic machinery. The gene or sequence may include nonfunctional sequences or sequences with no known function. A host cell that receives and expresses introduced DNA or RNA has been "transformed" and is a "transformant" or a "clone." The DNA or RNA introduced to a host cell can come from any source, including cells of the same genus or species as the host cell, or cells of a different genus or species.

The terms "vector", "cloning vector" and "expression vector" mean the vehicle by which a DNA or RNA sequence (*e.g.* a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (*e.g.* transcription and translation) of the introduced

sequence. Vectors include plasmids, phages, viruses, etc.; they are discussed in greater detail below.

Vectors typically comprise the DNA of a transmissible agent, into which heterologous DNA is inserted. A common way to insert one segment of DNA into another segment of DNA involves the use of enzymes called restriction enzymes that cleave DNA at specific sites (specific groups of nucleotides) called restriction sites. A "cassette" refers to a DNA coding sequence or segment of DNA that codes for an expression product that can be inserted into a vector at defined restriction sites. The cassette restriction sites are designed to ensure insertion of the cassette in the proper reading frame. Generally, foreign DNA is inserted at one or more restriction sites of the vector DNA, and then is carried by the vector into a host cell along with the transmissible vector DNA. A segment or sequence of DNA having inserted or added DNA, such as an expression vector, can also be called a "DNA construct." A common type of vector is a "plasmid", which generally is a self-contained molecule of double-stranded DNA, usually of bacterial origin, that can readily accept additional (foreign) DNA and which can readily introduced into a suitable host cell. A plasmid vector often contains coding DNA and promoter DNA and has one or more restriction sites suitable for inserting foreign DNA. Promoter DNA is a DNA sequence which initiates, regulates, or otherwise mediates or controls the expression of the coding DNA. Promoter DNA and coding DNA may be from the same gene or from different genes, and may be from the same or different organisms. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. Non-limiting examples include pKK plasmids (Clonetech), pUC plasmids, pET plasmids (Novagen, Inc., Madison, WI), pRSET or pREP plasmids (Invitrogen, San Diego, CA), or pMAL plasmids (New England Biolabs, Beverly, MA), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, *e.g.* antibiotic resistance, and one or more expression cassettes.

The term "host cell" means any cell of any organism that is selected, modified, transformed, grown, or used or manipulated in any way, for the production of a substance by the cell, for example the expression by the cell of a gene, a DNA or RNA sequence, a protein or an enzyme. Host cells can further be used for screening or other assays, as described *infra*. In a preferred aspect, a host cell of the invention is an actinomycete, preferably of the genus *Streptomyces* (*e.g.*, a host cell as described in Ziermann and Betlach, BioTechniques, 1999,

26:106) or alternatively *Micromonospera*. Additional examples include, but are not limited to, the strains *S. pristinaespiralis* (ATCC 25486), *S.antibioticus* (DSM 40868), *S. bikiniensis* (ATCC 11062), *S. parvulus* (ATCC 12434), *S. glauescens* (ETH 22794), *S.actuosus* (ATCC 25421), *S. coelicolor* (A3(2)), *S. ambofaciens*, *S. lividans*, *S. griseofuscus*, *S. limosus*, and the like (*see also* Smokvina et al., Proceedings, 1:403-407).

The term "expression system" means a host cell and compatible vector under suitable conditions, *e.g.,* for the expression of a protein coded for by foreign DNA carried by the vector and introduced to the host cell. Common expression systems include *E. coli* host cells and plasmid vectors, although the actinomycte host cell expression systems are preferred for biosynthesis of everninomicin and related products.

The term "heterologous" refers to a combination of elements not naturally occurring. For example, heterologous DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. A heterologous gene is a gene in which the regulatory control sequences are not found naturally in association with the coding sequence. In the context of the present invention, an *EV biosynthetic enzyme* gene is heterologous to the vector DNA in which it is inserted for cloning or expression, and it is heterologous to a host cell containing such a vector, in which it is expressed, *e.g.*, a K562 cell.

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (*see* Sambrook *et al.*, *supra*). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a T_m (melting temperature) of 55 °C, can be used, *e.g.*, 5x SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5x SSC, 0.5% SDS). Moderate stringency hybridization conditions correspond to a higher T_m, *e.g.*, 40% formamide, with 5x or 6x SCC. High stringency hybridization conditions correspond to the highest T_m, *e.g.*, 50% formamide, 5x or 6x SCC. SCC is a 0.15M NaCl, 0.015M Na-citrate. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or

homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (*see* Sambrook *et al.*, *supra*, 9.50-9.51). For hybridization with shorter nucleic acids, *i.e.*, oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (*see* Sambrook *et al.*, *supra*, 11.7-11.8). A minimum length for a hybridizable nucleic acid is at least about 10 nucleotides; preferably at least about 15 nucleotides; and more preferably the length is at least about 20 nucleotides.

In a specific embodiment, the term "standard hybridization conditions" refers to a T_m of 55 °C, and utilizes conditions as set forth above. In a preferred embodiment, the T_m is 60 °C; in a more preferred embodiment, the T_m is 65 °C. In a specific embodiment, "high stringency" refers to hybridization and/or washing conditions at 68 °C in 0.2XSSC, at 42 °C in 50 % formamide, 4XSSC, or under conditions that afford levels of hybridization equivalent to those observed under either of these two conditions.

As used herein, the term "oligonucleotide" refers to a nucleic acid, generally of at least 10, preferably at least 15, and more preferably at least 20 nucleotides, preferably no more than 100 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule encoding a gene, mRNA, cDNA, or other nucleic acid of interest. Oligonucleotides can be labeled, *e.g.*, with ^{32}P-nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. In one embodiment, a labeled oligonucleotide can be used as a probe to detect the presence of a nucleic acid. In another embodiment, oligonucleotides (one or both of which may be labeled) can be used as PCR primers, either for cloning full length or a fragment of EV biosynthetic enzyme, or to detect the presence of nucleic acids encoding EV biosynthetic enzyme. In a further embodiment, an oligonucleotide of the invention can form a triple helix with a EV biosynthetic enzyme DNA molecule. Generally, oligonucleotides are prepared synthetically, preferably on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thioester bonds, etc.

EV Biosynthetic Pathway Nucleic Acids

A gene encoding EV biosynthetic enzyme can be isolated from any everninomicin-producing *Micromonospora* source. Methods for obtaining EV biosynthetic enzyme gene are well known in the art, as described above (*see, e.g.*, Sambrook *et al.*, 1989, *supra*). The DNA may be obtained by standard procedures known in the art from cloned DNA, by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA (*e.g.*, DNA having a sequence as deposited with the ATCC and accorded accession no. 39149), or fragments thereof, purified from the desired cell (*see*, for example, Sambrook *et al.*, 1989, *supra*; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II). Whatever the source, the gene can be molecularly cloned into a suitable vector for propagation of the gene. Identification of the specific DNA fragment containing the desired *EV biosynthetic enzyme* gene may be accomplished in a number of ways. For example, a portion of an *EV biosynthetic enzyme* gene exemplified *infra* can be purified and labeled to prepare a labeled probe, and the generated DNA may be screened by nucleic acid hybridization to the labeled probe (Benton and Davis, Science, 1977, 196:180; Grunstein and Hogness, Proc. Natl. Acad. Sci. U.S.A., 1975, 72:3961). Those DNA fragments with substantial homology to the probe, such as an allelic variant from another species, will hybridize. In a specific embodiment, highest stringency hybridization conditions are used to identify a homologous *EV biosynthetic enzyme* gene.

Further selection can be carried out on the basis of the properties of the gene, *e.g.*, if the gene encodes a protein product having the isoelectric, electrophoretic, amino acid composition, partial or complete amino acid sequence, antibody binding activity, or ligand binding profile of EV biosynthetic enzyme protein as disclosed herein. Thus, the presence of the gene may be detected by assays based on the physical, chemical, immunological, or functional properties of its expressed product.

Other DNA sequences which encode substantially the same amino acid sequence as an *EV biosynthetic enzyme* gene may be used in the practice of the present invention. These include but are not limited to allelic variants, species variants, sequence conservative variants, and functional variants.

The genes encoding EV biosynthetic enzyme derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned EV biosynthetic enzyme gene sequence can be modified by any of numerous strategies known in the art

(Sambrook *et al.*, 1989, *supra*). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of EV biosynthetic enzyme, care should be taken to ensure that the modified gene remains within the same translational reading frame as the *EV biosynthetic enzyme* gene, uninterrupted by translational stop signals, in the gene region where the desired activity is encoded, unless the gene will be used to knock-out or disrupt an endogenous EV biosynthetic enzyme.

Additionally, the EV biosynthetic enzyme-encoding nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Such modifications can also be made to introduce restriction sites and facilitate cloning the *EV biosynthetic enzyme* gene into an expression vector. Any technique for mutagenesis known in the art can be used, including but not limited to, *in vitro* site-directed mutagenesis (Hutchinson, C., *et al.*, J. Biol. Chem., 1978, 253:6551; Zoller and Smith, DNA, 1984, 3:479-488; Oliphant *et al.*, Gene 1986, 44:177; Hutchinson *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1986, 83:710), use of TAB™ linkers (Pharmacia), etc. PCR techniques are preferred for site directed mutagenesis (see Higuchi, "Using PCR to Engineer DNA", in *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, ed., Stockton Press, 1989, Chapter 6, pp. 61-70).

The identified and isolated gene can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Examples of vectors include, but are not limited to, *E. coli*, bacteriophages such as lambda derivatives, or plasmids such as pBR322 derivatives or pUC plasmid derivatives, *e.g.*, pGEX vectors, pmal-c, pFLAG, etc. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. Finally, the vector may include a fusion polypeptide sequence such that

the construct with the EV biosynthetic enzyme encodes a chimeric protein, such as a poly-histidine tag, FLAG tag, myc epitope tag, or some other such sequence for ease in purification.

Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, etc., so that many copies of the gene sequence are generated. Preferably, the cloned gene is contained on a shuttle vector plasmid, which provides for expansion in a cloning cell, *e.g.*, *E. coli*, and facile purification for subsequent insertion into an appropriate expression cell line, if such is desired.

Expression of EV Biosynthetic Enzyme Polypeptides

The nucleotide sequence coding for EV biosynthetic enzyme, or antigenic fragment, derivative or analog thereof, or a functionally active derivative, including a chimeric protein, thereof, can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. Thus, a nucleic acid encoding EV biosynthetic enzyme of the invention can be operationally associated with a promoter in an expression vector of the invention. Such vectors can be used to express functional or functionally inactivated EV biosynthetic enzyme polypeptides.

The necessary transcriptional and translational signals can be provided on a recombinant expression vector.

Expression of EV biosynthetic enzyme protein may be controlled by any promoter/enhancer element known in the art, but these regulatory elements must be functional in the host selected for expression. Promoters which may be used to control EV biosynthetic enzyme gene expression include, but are not limited to, prokaryotic expression vectors such as the β-lactamase promoter (Villa-Komaroff, *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1978, 75:3727-3731), or the *tac* promoter (DeBoer, *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1983, 80:21-25; see also "Useful proteins from recombinant bacteria" in Scientific American, 242:74-94, 1980). Among regulable promoters which can be used in the context of the present invention, mention may be made more especially of any regulable promoter which is functional in actinomycetes. These can comprise promoters induced specifically by an agent introduced into to the culture medium, such as, for example, the thiostrepton-inducible promoter tipA (Murakami *et al.*, J.Bact., 1989, 171:1459), or thermoinducible promoters such as that of the groEL genes, for example (Mazodier *et al.*, J.Bact., 1991, 173:7382). They can also comprise

an actinomycetes promoter which is specifically active in the late phases of the proliferation cycle of actinomycetes, such as, for example, certain promoters of genes of the secondary metabolism (genes for the production of antibiotics, in particular).

Soluble forms of the protein can be obtained by collecting culture fluid, or solubilizing inclusion bodies, *e.g.*, by treatment with detergent, and if desired sonication or other mechanical processes, as described above. The solubilized or soluble protein can be isolated using various techniques, such as polyacrylamide gel electrophoresis (PAGE), isoelectric focusing, 2-dimensional gel electrophoresis, chromatography (*e.g.*, ion exchange, affinity, immunoaffinity, and sizing column chromatography), centrifugation, differential solubility, immunoprecipitation, or by any other standard technique for the purification of proteins.

Antibodies to EV Biosynthetic Enzymes

According to the invention, any EV biosynthetic enzyme polypeptide produced recombinantly or by chemical synthesis, and fragments or other derivatives or analogs thereof, including fusion proteins, may be used as an immunogen to generate antibodies that recognize the EV biosynthetic enzyme polypeptide. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. The anti-EV biosynthetic enzyme antibodies of the invention may be cross reactive, *e.g.*, they may recognize EV biosynthetic enzyme from different species. Polyclonal antibodies have greater likelihood of cross reactivity. Alternatively, an antibody of the invention may be specific for a single form of EV biosynthetic enzyme, such as murine EV biosynthetic enzyme. Preferably, such an antibody is specific for human EV biosynthetic enzyme.

Various procedures known in the art may be used for the production of polyclonal antibodies to EV biosynthetic enzyme polypeptide or derivative or analog thereof. For the production of antibody, various host animals can be immunized by injection with the EV biosynthetic enzyme polypeptide, or a derivative (*e.g.*, fragment or fusion protein) thereof, including but not limited to rabbits, mice, rats, sheep, goats, etc. In one embodiment, the EV biosynthetic enzyme polypeptide or fragment thereof can be conjugated to an immunogenic carrier, *e.g.*, bovine serum albumin (BSA) or keyhole limpet hemocyanin (KLH). Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and

potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum*.

For preparation of monoclonal antibodies directed toward the EV biosynthetic enzyme polypeptide, or fragment, analog, or derivative thereof, any technique that provides for the production of antibody molecules by continuous cell lines in culture may be used. These include but are not limited to the hybridoma technique originally developed by Kohler and Milstein (Nature, 1975, 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today, 1983, 4:72; Cote *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1983, 80:2026-2030), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., 1985, pp. 77-96).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent Nos. 5,476,786 and 5,132,405 to Huston; U.S. Patent 4,946,778) can be adapted to produce EV biosynthetic enzyme polypeptide-specific single chain antibodies. Indeed, these genes can be delivered for expression *in vivo*. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse *et al.*, Science, 1989, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for an EV biosynthetic enzyme polypeptide, or its derivatives, or analogs.

Antibody fragments which contain the idiotype of the antibody molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.*, radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by detecting a label on the

primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. For example, to select antibodies which recognize a specific epitope of an EV biosynthetic enzyme polypeptide, one may assay generated hybridomas for a product which binds to an EV biosynthetic enzyme polypeptide fragment containing such epitope. For selection of an antibody specific to an EV biosynthetic enzyme polypeptide from a particular species of animal, one can select on the basis of positive binding with EV biosynthetic enzyme polypeptide expressed by or isolated from cells of that species of animal.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the EV biosynthetic enzyme polypeptide, *e.g.*, for Western blotting, imaging EV biosynthetic enzyme polypeptide *in situ*, measuring levels thereof in appropriate physiological samples, etc. using any of the detection techniques mentioned above or known in the art.

In a specific embodiment, antibodies that agonize or antagonize the activity of EV biosynthetic enzyme polypeptide can be generated. Such antibodies can be tested using the assays described *infra* for identifying ligands.

Techniques of isolating bacterial DNA are readily available and well known in the art. Any such techniques can be employed in this invention. In particular DNA from these deposited cultures can be isolated as follows. Lyophils of *E. coli* XL1-Blue/pSPRX272, *E. coli* XL1-Blue/pSPRX2262, *E. coli* XL1-Blue/pSPR192, *E. coli* XL1-Blue/pSPRX210 or *E. coli* XL1-Blue/pSPRX256 are plated onto L-agar (10 g tryptone, 10 g NaCl, 5 g yeast extract, and 15 g agar per liter) plates containing 100 µg/ml ampicillin to obtain a single colony isolate of the strain. This colony is used to inoculate about 500 ml of L-broth (10 g tryptone, 10 g NaCl, 5 g yeast extract per liter) containing 100 µg/ml apramycin, and the resulting culture is incubated at 37 °C with aeration until the cells reach stationary phase. Cosmid DNA can be obtained from the cells in accordance with procedures known in the art (*see*, *e.g.*, Rao *et al.*, Methods in Enzymology, 1987, 153:166).

DNA of the current invention can be sequenced using any known techniques in the art such as the dideoxynucleotide chain-termination method (Sanger *et al.*, Proc. Natl. Acad. Sci., 1977, 74:5463) with either radioisotopic or fluorescent labels. Double-stranded, supercoiled

DNA can be used directly for templates in sequence reactions with sequence-specific oligonucleotide primers. Alternatively, fragments can be used to prepare libraries of either random, overlapping sequences in the bacteriophage M13 or nested, overlapping deletions in a plasmid vector. Individual recombinant DNA subclones are then sequenced with vector-specific oligonucleotide primers. Radioactive reaction products are electrophoresed on denaturing polyacrylamide gels and analyzed by autoradiography.

Fluorescently labeled reaction products are electrophoresed and analyzed on Applied Biosystems (ABI Division, Perkin Elmer, Foster City, Calif. 94404) model 370A and 373A or Dupont (Wilmington, Del.) Genesis DNA sequencers. Sequence data are assembled and edited using Genetic Center Group (GCG, Madison, Wis.) programs GelAssemble and Seqed or the ABI model 670 Inherit Sequence Analysis system and the AutoAssembler and SeqEd programs.

Polypeptides corresponding to a domain, a submodule, a module, a synthesis unit (SU), or an open reading frame can be produced by transforming a host cell such as bacteria, yeast, or eukaryotic cell-expression system with the cDNA sequence in a recombinant DNA vector. It is well within one skilled in the art to choose among host cells and numerous recombinant DNA expression vectors to practice the instant invention. Multifunctional polypeptides of polyketide everninomicin synthase can be extracted from everninomicin-producing bacteria such as Streptomyces ambofaciens or translated in a cell-free in vitro translation system. In addition, the techniques of synthetic chemistry can be employed to synthesize some of the polypeptides mentioned above.

Procedures and techniques for isolation and purification of proteins produced in recombinant host cells are known in the art. See, for example, Roberts *et al.*, Eur. J. Biochem., 1993, 214: 305-311 and Caffrey *et al.*, FEBS, 1992, 304:225-228 for detailed description of polyketide synthase purification in bacteria. To achieve a homogeneous preparation of a polypeptide, proteins in the crude cell extract can be separated by size and/or charge through different columns well known in the art once or several times. In particular the crude cell extract can be applied to various cellulose columns commercially available such as DEAE-cellulose columns. Subsequently the bound proteins can be eluted and the fractions can be tested for the presence of the polyketide everninomicin synthase or engineered derivative protein. Techniques for detecting the target protein are readily available in the art. Any such techniques can be employed for this invention.

In particular the fractions can be analyzed on Western blot using antibodies raised against a portion or portions of such polyketide everninomicin synthase proteins. The fractions containing the polyketide everninomicin synthase protein can be pooled and further purified by passing through more columns well known in the art such as applying the pooled fractions to a gel filtration column. When visualized on SDS-PAGE gels homogeneous preparations contain a single band and are substantially free of other proteins.

Actinomycetes are prolific producers of secondary metabolites with antimicrobial and antifungal activity and represent a significant source of active compounds for pharmaceuticals. The genus *Streptomyces* produces a wide variety of secondary metabolites including antitumor, antifungal, and antimicrobial agents. The biosynthesis of these compounds has been shown to be directed by large multi-functional proteins or a number of proteins each catalyzing specific steps in the biosynthesis of the secondary metabolite (REF- Biotechnology of AB etc.) The genes encoding actinomycete secondary metabolite biosynthesis have been found to be clustered on contiguous segments of each producing organisms genome (Strohl, William R., 1997, Biotechnology of Antibiotics, 2nd Ed., Marcel Dekker, Inc., New York, NY). This makes it feasible for complete pathways to be cloned, analyzed, genetically manipulated and expressed in surrogate hosts.

Components of The Everninomicin Biosynthethic Pathway

Orsellinic Acid Biosynthesis

The term "polyketide" refers to a class of molecules produced through the successive condensation of small carboxylic acids. This diverse group includes plant flavonoids, fungal aflatoxins, and hundreds of compounds of different structures that exhibit antibacterial, antifungal, antitumor, and anthelmintic properties. Some polyketides produced by fungi and bacteria are associated with sporulation or other developmental pathways; others do not yet have an ascribed function. Some polyketides have more than one pharmacological effect. The diversity of polyketide structures reflects the wide variety of their biological properties. Many cyclized polyketides undergo glycosidation at one or more sites, and virtually all are modified during their synthesis through hydroxylation, reduction, epoxidation, etc.

For the purposes of the present invention, "polyketide" refers to the orsellenic acid moiety in everninomicin. Thus, the invention provides, in particular, the DNA sequence encoding the polyketide synthase responsible for biosynthesis of this orsellinic acid moiety of everninomicin, *i.e.*, the everninomicin orsellinic acid synthetase. The everninomicin orsellinic

acid synthase DNA sequence, which defines the orsellinic synthase gene cluster, directs biosynthesis of the orsellinic acid polyketide by encoding the various distinct activities of orsellinic synthase. The skilled artisan recognizes, however, that the everninomicin orsellinic synthase genes are useful in the production of other polyketides, *e.g.*, by recapitulating all or part of this component of the biosynthetic pathway, or by modulating biosynthetic pathways (*see*, the discussion about combinatorial biosynthesis, *infra*).

The gene cluster for orsellinic synthase, like other Type I polyketide biosynthetic synthase genes whose organization has been elucidated, is characterized by the presence of an ORF encoding a multi-functional protein which contains separate, active sites for condensation of acyl groups as defined above. The map of the orsellinic synthase gene derived from *Micromonospora carbonacea var. africana* is shown in Figure 3. The accompanying synthetic pathway and the specific carboxylic acid substrates that are used for each condensation of orsellinic acid synthesis are indicated in Figure 4.

Polyketides are complex secondary metabolites synthesized from the condensation of acetyl-coenzyme A (CoA) or related acyl-CoAs by polyketide synthetase enzymes. Other acyl groups forming the acyl-CoA include malonyl, proponyl, and butyryl. Condensation of extender units requires the action of β-ketoacyl ACP synthetase, acetyltransferase and acyl carrier protein enzymatic sites. Each module processes one condensation step and typically requires several activities accomplished by several active sites including acyl carrier protein (ACP), β-ketosynthase (KS), and acyltransferase (AT). The specific gene products identified with orsellinic biosynthesis are listed in Table 2.

Table 2. Orsellinic Acid Biosynthetic Gene Products

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evrF	21,064...22,542	19	non-heme oxygenase/halogenase addition
evrI	25,550...26,626	22	acyl starter unit
evrJ	26,685...30,479	23	orsellinic acid synthase/6-methylsalicilic acid synthase
evbD	56,961...58,709	47	acyl-CoA carboxylase
evbQ	74,707...76,290*	62	Methylmalonyl-CoA mutase

Polyketide synthetases are classified as either iterative Type I, iterative Type II or modular polyketide sythetases. Iterative Type I synthetases resemble the multifunctional fatty acid synthases from animals and are composed of multifunctional proteins with separate

protein domains encoding each active sites. This is exemplified by the actinomycete *S. erythrea* polyketide synthetase for the biosynthesis of erythromycin, the *Streptomyces viridochromogenes* Tu57 AviM synthesis of orsellinic acid and the *Penicillium patulum* polyketide synthase for 6-methylsalicylic acid (Hutchinson *et al.*, Annual Review of Microbiology, 1995, 49:201-238; Gaisser *et al.*, Journal of Bacteriology, 1997, 179:6271-6278; Beck *et al.*, European Journal of Biochemistry, 1990, 192:487-498). Iterative type II synthetases have seperate proteins for each active site. These are exemplified by the polyketide synthetases from *S. coelicolor, S. violaceoruber* and *S. glaucescens* synthesizing the aromatic polyketides actinorhodin, granaticin and tetracenomycin respectively (Hopwood, *et al.*, Annual Review of Microbiology 1990, 24:37-66). The modular polyketide synthetases are large proteins that contain several domains with each domain containing several active sites. An example of a modular polyketide synthetase is the 6-deoxyerythronolide B synthetase from *Saccharopolyspora erythraea*. Recent reviews of polyketides and polyketide synthetases elaborate on these pathways (Hopwood, *et al.*, Annual Review of Microbiology, 1990, 24:37-66; Hutchinson *et al.*, Annual Review of Microbiology, 1995, 49:201-238).

Although not wishing to be bound to any particular theory or technical explanation, a sequence similarity exists among domain boundaries in various polyketide synthase genes. Thus, one skilled in the art is able to predict the domain boundaries of newly discovered polyketide synthase genes based on the sequence information of known polyketide synthase genes. In particular, the boundaries of submodules, domains, and open reading frames in the instant application are predicted based on sequence information disclosed in this application and the locations of the domain boundaries of the everninomicin synthase (Donadio *et al.*, GENE, 1992, 111:51-60). Furthermore, the genetic organization of the everninomicin synthase gene cluster appears to correspond to the order of the reactions required to complete synthesis of everninomicin. This means that the polyketide synthase DNA sequence can be manipulated to generate predictable alterations in the final everninomicin product.

Acyl Precursor Formation

EvrJ (orsellinic acid synthetase) requires one acetyl-CoA starter and three malonyl-CoA extender units to synthesize orsellinic acid. The acetyl-CoA and malonyl-CoA units most likely are derived from glycolysis and fatty acid biosynthesis (Tang L, *et al.*, Ann. N Y Acad. Sci., 1994, 721:105-16). The malonyl-CoA can also be derived from acetyl-CoA by carboxylation by acetylCoA carboxylase,(Scott Eagleson, Concise Encyclopedia of Biochemistry, 2nd Ed., Walter de Gruyler; Berlin, 1988). The *M. carbonacea* EV region

contains an *evbD* which has strong homology to know acetyl-CoA carboxylases. Thus evbD is responsible for the synthesis of the malonyl-CoA unit required for orsellinic acid biosynthesisas shown in Figure 4.

Sugar Biosynthetic Products and Glycosyltransferases

Glycosyl groups (6-deoxysugars) are synthesized by a common mechanism involving hexose-1-P nucleotidyl-transferase, dTDP-D-glucose synthetase and dTDP-D-glucose 4,6-dehydratase. L-deoxysugars are synthesized by the action of a NDP-4-keto-6-deoxyhexose 3,5-epimerase. Deoxysugars can be modified by deoxygenations, transaminations, methylations and isomerization or epimerizations prior to covalent attachment by a glycosytransferase.

Biosynthesis of the sugars (*see* Liu and Thorson, Annu. Rev. Microbiol., 1994, 48:223) that are attached to the orsellinic acid/polyketide, and the enzymes that mediate attachment of the sugars, are also key elements of the everninomicin biosynthetic pathway. Genes encoding such sugar biosynthetic enzymes and glycosyltransferases are typically found in the biosynthetic pathway locus (*see* Summers *et al.*, Microbiology, 1997, 143:3251). The genes identified from the EV biosynthetic locus are listed in Tables 3 and 4.

Table 3. Sugar Biosynthetic Gene Products

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evdA	132...1382*	2	hydroxylase
evdB	1490...2611*	3	hexose aminotransferase
evdC	2622...3860*	4	oxidase (flavoprotein)
evdE	5309...6235	6	hexose dehydratase
evdI	9463...10,224*	10	hydrolase
evdK	11,208...12,455	12	hexose dehydratase or epimerase
evrA	14,410...15,363*	14	hexose epimerase
evrB	15,380...16,414*	15	hexose oxidoreductase
evrC	16,419...17,873*	16	hexose dehydratase
evrD	17,870...18,934*	17	GDP-mannose 4,6-dehydratase
evrV	41,679...42,707*	35	dTDP-glucose epimerase
evrW	42,810...43,799*	36	dTDP-glucose dehydratase
evrX	43,799...44,866	37	dTDP-glucose synthetase

evbS	78,791...80,521	64	phosphomannomutase
evbU	83,280...83,888	66	glucose-6-phosphate 1-dehydrogenase
ORF9	8254...9318	104	oxidoreductase
ORF11	10,584...11,585	106	deoxyhexose ketoreductase

Table 4. Glycosyltransferases

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evdD	4143...5312	5	DNTP-hexose glycosyltransferase
evdF	6232...7275	7	DNTP-hexose glycosyltransferase
evdH	8342...9364	9	DNTP-hexose glycosyltransferase
evdL	12,108...13,022*	13	DNTP-hexose glycosyltransferase
evrS	38,892...40,163*	32	DNTP-hexose glycosyltransferase

These genes are important targets for modulation. They are likely to be bottleneck genes, and thus increased expression using an exogenous or integrating vector can increase the yield of everninomicin (or its analog). Alternatively, knocking out these genes may result in complete elimination of everninomicin biosynthesis.

Tailoring Enzymes

Various types of EV biosynthetic enzymes fall into the tailoring enyzme category. These are listed in Table 5. Increasing or decreasing expression of these enzymes permits production of everninomicin analogs. Moreover, expression of these enzymes in other actinomycetes permits production of novel secondary metabolites by the action of the everninomicin tailoring enzymes on these metabolites.

Table 5. Tailoring Gene Products

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evrG	22,748...24,172	20	oxidase
evrL	31,941...32,882*	25	heme biosynthesis
evrM	33,167...34,405*	26	p450 hydroxylase
evrN	34,449...35,210*	27	methyl transferase
evrQ	36,998...38,026*	30	oxidoreductase/heat stress protein
evrT	40,216...40,890	33	L-proline hydroxylase
evrU	40,887...41,576	34	methyltransferase
evbA	53,554...54,207	43	o-methyltransferase
evbE	58,873...60,312	48	IMP dehydrogenase
evbI	66,469...67,872*	54	lipoamide dehydrogenase
evbL	69,610...70,359*	57	acetyltransferase/phosphotransferase
evbX	85,909...87,342	69	aldehyde dehydrogenase
exbY	87,422...88159	70	aldehyde dehydrogenase
evcB	89,817...91,067	73	cytochrome D oxidase subunit I
evcC	91,078...92,085	74	cytochrome D oxidase subunit II

Regulatory Products: Serine-Threonine Kinases

Protein serine (Ser), threonine (Thr), and tyrosine (Tyr) kinases play essential roles in signal transduction in organisms ranging from yeast to mammals, where they regulate a diverse cellular activities. Genes that encode eukaryotic-type protein kinases have also been identified in different bacterial species, suggesting that such enzymes are also widespread in prokaryotes. Although many of them have yet to be fully characterized, several studies indicate that eukaryotic-type protein kinases play important roles in regulating cellular activities of these bacteria, such as cell differentiation and secondary metabolism (Cheng-Cai, Molecular Microbiology, 1996, 20:9-15). Examples that have been studied include the *pknD* Ser/Thr kinase from *Anabaena sp. PCC7120*, which is involved in nitrogen metabolism control (Zhang *et al.*, Molecular and General Genetics, 1998, 258:26-33); the *pkn9* Ser/Thr kinase from *Myxococcus xanthus*, which is involved in development of fruiting bodies (Hanlon *et al.*, Molecular Microbiology, 1997, 23:459-71); and the *afsK* Ser/Thr kinase from *Streptomyces coelicolor*, which is involved in aerial myceliaum formation (Ueda *et al.*, Gene, 1996, 169:91-95). These genes from the EV biosynthetic locus are listed in Table 6.

Table 6. Regulatory Gene Products

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evrR	38,072...38,566	31	hexaheme nitrite reductase regulator/ methyltransferase
evsA	47,156...49,234*	40	serine-threonine kinase
evbF	60,472...61,029*	49	
evbF2	61,610...62,069	51	
evbK	68,529...69,494*	56	protease synthase/sporulation regulator
evbR	76,622...78,712	63	protein serine-threonine kinase (eukaryotic type)
evcJ	100,733...101,326*	81	ATP/GTP binding protein
ORF1	189...1064*	96	Transcriptional regulator biotinylation
ORF4	3776...4276*	99	ECF sigma factor

The evsA and evbR proteins within the everninomicin cluster have a high degree of homology to Ser/Thr kinases and may play a role in regulating the expression of the pathway. Manipulation of the evsA and evbR proteins could enhance the expression and yield of everninomicin from *M. carbonacea* by providing positive signals for biosynthesis. Thus, these genes are preferred elements in a vector to enhance the efficiency of everninomicin biosynthesis.

Resistance Mechanisms

Actinomycetes utilize a variety or mechanisms to confer resistance to secondary metabolites they produce. These include membrane pumps, rRNA methylases, O-phosphorylation, N-acetylation, and production of resistant target proteins (Cundliffe, Annual Review of Microbiology, 1989, 43:207-33). The genes from the EV biosynthetic locus that have this function are listed in Table 7.

Table 7. Resistance Mechanism Genes

Gene Product	CDS	SEQ ID No.	Enzymatic Function
evrE	19,374...20,906	18	multidrug eflux transporter
evrY	45,014...45,760*	38	dehalogenase
evrZ	45,962...46,714*	39	muramidase/lysozyme
evbB	54,362...55,117*	44	membrane pump

evbC	55,135...56,094*	45	membrane pump
evbC2	56,184...56,813*	46	ankrylin-like
evbG	62,122...63,795	52	ABC transporter
evbH	63,891...65,828	53	ABC transporter
evcD	92,148...93,833	75	ABC transporter
evcE	93,830...95,671	76	ABC transporter
evrMR	107,653...108,615	87	23S rRNA methylase
evrMR12	108,635...109,216	88	
ORF6	5392...6147*	101	rRNA methyltransferase

Multi-drug transporters are membrane proteins that are able to expel a broad range of toxic molecules from the microbial cells. These multidrug transporters belong to the ATP-binding cassette (ABC) family of transport proteins that utilize the energy of ATP hydrolysis for activity. In microorganisms, multidrug transporters play an important role in conferring antibiotic resistance on pathogens, and in actinomycetes confer resistance to the antibiotic secondary metabolites produced by these organisms themselves (Fath *et al.*, Microbial Reviews, 1993, 57:995-1017). A second class of membrane transporters that are found in actinomycetes include MDR (multiple drug resistance) type pumps found in eukaryotes (Guilfoile *et al.*, Proc. Natl. Acad. Sci. USA, 1991, 88:8553-8557). The EV cluster contains evbB and evbC, which are homologouse to the ATP-binding cassette (ABC) family of transport proteins and specifically to the mithramycin resistance pump from *Streptomyces argillaceus* (Fernandez *et al.*, Molecular and General Genetics, 1996, 251:692-698). In addition the EV cluster contains evrE, an MDR type pump with homology to the *Streptomyces peucetius* drrA MDR type pump that confers resistance to daunorubicin. Ribosomal methylases have also been found to confer resistance to producing organisms. The tlrB 23S rRNA methylase from *Streptomyces fradiae* and the myrA 23S rRNA methylase from *Micromonospora griseorubida* have been found to confer resistance to tylosin and mycinamicin respectively.

The EV cluster also contains evrMR, a 23 RNA methylase with (loc.) homology to both tlrB and myrA.

The EV pathway also contains *evrZ*, a gene with homology to muramidases. Muramidases (lysozyme) cleave β1,4 linkages between *N*-acetylglucosamine and

N-acetylmuramic acid (Scott and Eagleson, Concise Encyclopedia Biochemistry, 2nd Ed., Walter de Gruyter: Berlin, 1988 p. 353). Thus, evrZ may inactivate everninomicin by cleavage within the glycosyl bonds.

Increased levels of expression of one or more of these resistance genes is expected to enhance the efficiency of everninomicin biosynthesis in an enhanced biosynthetic system by reducing toxicity to the host cell.

Furthermore, these resistance genes are good candidates for use as positive selection markers in recombinant systems. By including an everninomicin resistance gene in a vector, a host cell successfully transformed with the vector will demonstrate everninomicin resistance. Thus, everninomicin becomes a useful tool for selecting transformed host cells.

Biosynthetic Production And Modification of Everninomicins

There are a number of uses for the cloned *Micromonospora carbonacea* EV cluster DNA. The cloned genes can be used to improve the yields of everninomicins and to produce novel everninomicins. Improved yields can be obtained by introduction of a second copies of genes for enzymes that are rate limiting in the pathway ("bottleneck genes"). This can be accomplished by cloning genes onto vectors, preferably integrating vectors, then obtaining integrants in the chromosome. Alternatively, a rate limiting enzyme gene can be modified by associating it with a strongly expressing promoter sequence and then integrating this construct into the chromosome. Manipulation of regulatory proteins including the Ser/Thr kinases can enhance yields by obtaining mutants that express EV pathway genes at higher levels than parental organisms.

Novel everninomicins can be produced by using cloned fragments to disrupt steps in the biosynthesis of everninomicin. Disruptions can lead to the accumulation of precursors or "shunt" products. To generate disruptions, DNA fragments of internal segments of genes (lacking 5' and 3' sequences) can be cloned into insertion vectors. These constructs can be introduced into the parental organism and homologous recombinants selected for that result in two copies of the gene in the chromosome. One copy lacks 3' sequences and the second copy lacks upstream native promoter sequences and 5' sequences. Alternatively, DNA fragments of genes containing internal deletions or insertions can be cloned into gene replacement vectors. Recombinants can be obtained that contain internal deletions or insertions of genes, which results in a non-functional chromosome copy of the gene. Constructs that allow a frequency of recombination into the chromosome to obtain disruptions should contain fragments of

sufficient size for recombination to occur (300 to 600 bases). Modified everninomicins produced by disrupting the genes may be antibiotics themselves, or serve as substrates for further chemical modification, creating new semi-synthetic everninomicins with unique properties or spectra of activity.

Novel everninomicins can also be produced by mutagenesis of the cloned genes, and replacement of the mutated genes for their unmutated counterparts in the everninomicin producer. Mutagenesis may involve, for example, (1) manipulation of the orsellinic acid PKS TypeI gene by introduction of KR, DH or ER domains (*see*, Donidio *et al.*, 1993), *e.g.*, to yield a modified orsellenic acid nucleus; (2) manipulation of the glycosyltranferase to relax substrate or glycosyl specificity, *e.g.*, to yield everninomicin containing novel glycosyl groups or additional glycosyl groups; and/or (3) manipulation of glycosyl biosynthetic genes, *e.g.*, to yield novel glycosyl groups and everninomicin containing novel glycosyl groups.

The DNA from the everninomicin biosynthetic cluster can be used as a hybridization probe to identify homologous sequences. Thus, the DNA cloned here could be used to obtain uncloned regions flanking the region described here but not yet isolated. In addition DNA from the region cloned here may be useful in identification of non-identical but similar sequences in other organisms.

The modified strains provided by the invention may be cultivated to provide everninomicins using conventional protocols.

Genetic Manipulation of Actinomycetes

Protocols have been developed to genetically manipulate actinomycete genomes and biosynthetic pathways. These include *E. coli* actinomycete shuttle vectors, gene replacement systems, transformation protocols, transposon mutagenesis, insertional mutagenesis, integration systems and heterologous host expression. These techniques are reviewed in numerous articles (Baltz *et al.*, Trends Microbiol., 1998, 2:76-83, Hopwood *et al.*, Genetic Manipulation of Streptomyces: A Laboratory Manual, 1985; Wohlleben *et al.*, Acta Microbiol. Immunol. Hung, 1994, 41:381-9 [Review]).

The development of vectors for the genetic manipulation of actinomycetes began with the observation of plasmids in actinomycetes and the development of a transformation protocol of actinomycete protoplasts using polyethylene glycol (Bibb *et al.*, Nature, 1980, 284:526-31). Many standard molecular techniques for *Streptomyces* were developed by Hopwood and colleagues for *Streptomyces coelicolor* and *Streptomyces lividans* (Hopwood *et al.*, Genetic

Manipulation of Streptomyces: A Laboratory Manual, 1985). These techniques have been adapted and expanded to other actinomycetes.

Vectors incorporating antibiotic-resistance markers (AmR, ThR, SpR) that function in *Streptomyces spp.* and other features have allowed the development of vectors for (a) integration via homologous recombination between cloned DNA and the *Streptomyces spp.* chromosome, (b) autonomous replication, and (c) site-specific integration at the bacteriophage phiC31 attachment (att) site or pSAM2 attachment site, and (d) gene replacement vectors. Homologous recombination between the cloned DNA and the chromosome can be used to make insertional knockouts of specific genes. Autonomously replicating plasmids and integrating plasmids can be used to introduce heterologous genes into actinomycetes for complementation or expression studies.

Many actinomycetes contain restriction systems that limit the ability to transform organisms by protoplast transformation. More recent gene transfer procedures have been developed for introducing DNA into streptomycetes by conjugation from *Escherichia coli*. This employs a simple mating procedure for the conjugal transfer of vectors from *E. coli* to *Streptomyces spp.* that involves plating of the donor strain and either germinated spores or mycelial fragments of the recipient strain. Conjugal plasmids contain the 760-bp oriT fragment from the IncP plasmid, RK2 and are transferred by supplying transfer functions in trans by the *E. coli* donor strain. Other recent developments that increase the frequency of recombination of non-replicating plasmids into the recipient actinomycete chromosome include transformation of non-replicating plasmids into protoplasts using denatured plasmid DNA (Oh and Chater, J. Bacteriol., 1997, 179:122-7) and conjugation of non-replicating plasmids from a methyl minus strain of *E. coli*. (Smith *et al.*, FEMS Microbiol. Lett., 1997, 155:223-9).

Various strategies have been used to obtain gene replacements in streptomycetes, for the construction of mutations and the modification of biosynthetic pathways (Baltz *et al.*, 1998, *supra*; Hopwood *et al.*, *supra*; Wohllenben *et al.*, 1994, *supra*; Baltz and Hosted, TIBTECH, 1996, 14:245; Baltz, Curr. Op. Biotech., 1990, 1:12-20). These methods have typically employed a two or three step procedure that results in allelic exchange. Initial crossover events between a non-intergrating phage, non-replicating plasmid, or temperature sensitive plasmid and the streptomycete chromosome are selected for by antibiotic resistance. Subsequent recombination events that result in gene replacement can be detected by screening the progeny of the initial recombinants by PCR analysis, Southern analysis, appearance of an expected phenotype or screening for the loss of a resistance marker which had previously been

exchanged into the loci to be replaced. The last of these methods has been employed by Khosla *et al.*, Mol. Microbiol., 1992, 6:3237-49; Khosla *et al.*, J. Bacteriol., 1993, 175:2197-204, to successfully modify the polyketide biosynthetic route of *S. coelicolor*. The strategy employed by Khosla *et al.*, 1992, *supra*, also has the advantage of allowing placement of non-selectable and phenotypically silent alleles into chosen positions of the chromosome. Donadio *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1993, 90:7119-23 has also successfully reprogrammed the erythromycin pathway of *Saccharopolyspora erythrae* by gene replacement.

Non-replicating plasmids for gene replacement were initialy utilized by Hilleman *et al.*, Nucleic Acids Res., 1991, 19:727-31, who used a derivative of pDH5 to construct mutations in the phosphinothricin tripeptide biosynthetic pathway of *S.hygroscopicus*. Plasmid-integration events were obtained by thiostrepton selection, subsequent screening of the primary recombinants indicated that 4 of 100 isolates had undergone a double-crossover gene replacement.

Use of counterselectable or negative selection markers such as *rpsL* (confers streptomycin sensitivity) or *sacB* (confers sucrose sensitivity) have been widely employed in other microorganisms for selection of recombination that results in gene replacement. In *S. coelicolor*, Buttner utilized *glk* as a counterselectable marker in att minus phiC31 phage to select for recombination events to construct gene replacement mutants of three *S. coelicolor* RNA polymerase sigma factors (Buttner *et al.*, J. Bacteriol., 1990, 172:3367-78). Hosted has developed a gene replacement system utilizing the *rpsL* gene for counterselection (Hosted and Baltz, J. Bacteirol., 1997, 179:180-6).

The construction of recombinant streptomycete strains to produce hybrid secondary metabolites has been accomplished. Current procedures use recombinant DNA techniques to isolate and manipulate secondary metabolic pathways and to express these pathways in surrogate hosts such as *Streptomyces lividans*. Heterologous expression of diverse pathways, polyketide, oligopeptide and β-lactam biosynthetic pathways, has been achieved. Furthermore novel polyketide structures have been generated through the manipulation of polyketide genes forming chimeric pathways. Recently novel polyketide modules have been isolated from environmental sources using PCR amplification and expressed in Streptomyces to yield novel chemical structures (Strohl *et al.*, J. Industr. Microbiol., 1991, 7:163; Kim *et al.*, J. Bacteriol., 1995,77:1202; Ylihonko *et al.*, Microbiology, 1996, 142:1965).

Knowledge of the everninomicin synthase DNA sequence, its genetic organization, and the activities associated with particular open reading frames, modules, and submodules of the gene enables production of novel everninomicins that are not otherwise available. Modifications may be made to the DNA sequence that either alter the structure or sequence of addition of building blocks. The principles have already been described above. In addition, any product resulting from post-transcriptional or post-translational modification *in vivo* or *in vitro* based on the DNA sequence information disclosed here are meant to be encompassed by the present invention.

Combinatorial Biosynthesis

The EV biosynthetic enzymes described here are ideal candidates for combinatorial biosynthesis to generate libraries of orthomycins, particularly everninomicin analogs and homologs, for testing and drug discovery (*see* Altreuter and Clark, Curr. Op. Biotech., 1999, 10:130; Reynolds, Proc. Natl. Acad. Sci. USA, 1998, 95:112744). Moreover, unlike chemical synthesis, which may depend on the efficiency of a specific reaction to determine product yield, a biosynthetic system can be amplified and propagated to produce high yields of the desired product.

Actinomycetes are well known microbial biosynthetic factories, and have been modified to produce novel compounds by mutation of specific biosynthetic genes (*see* Hutchinson, Bio/Technology, 1994, 12:375; Piepersberg, Crit. Rev. Biotech., 1994, 14:251). In addition to mutagenisis *in situ*, rapid evolution by DNA shuffling, particularly with related genes from other species or from the EV biosynthetic locus itself, provides for more directed evolutionary mutagenesis (Stemmer, Nature, 1994, 370:389). This technique can be practiced, for example, by shuffling EV biosynthetic gene products with their closest homologs, as determined by BLAST (or some other homology algorithm) analysis. For example, gene shifting of two or more transferases can yield new enzymes with altered function. Similarly, sugar biosynthetic genes, orsellinic acid biosynthetic genes, and tailoring genes can be manipulated by the techniques of directed evolution, *e.g.*, gene shuffling, to produce mutants with novel enzymatic and synthetic function. Tailoring enzymes are particularly attractive targets for mutagenesis, since these will not affect synthesis of the core structure, but yield a variety of novel products.

An Integration Vector for *Micromonospera*

In a specific embodiment, the present invention relates to a new nucleic acid sequence, to vectors for its expression and to its use in fermentation processes in actinomycetes. This nucleic acid sequence encodes a *Micromonospera*, and particularly *M. carbonacea, var. africana,* att/int functions and thus permits development of an integrating vector. In a specific embodiment, the att/int functions has an amino acid sequence as depicted in SEQ ID NO:90. In a more specific embodiment, the integrase is encoded by a nucleic acid having a nucleotide sequence as depicted in SEQ ID NO:89 (Figure 7B). A preferred integrating plasmid is shown in Figure 7A.

Advantageously, the integrative vectors derived from this novel integrase also comprise a recombinant DNA sequence coding for a desired product, including but by no means limited to an EV biosynthetic gene. The product can be a peptide, polypeptide or protein of pharmaceutical or agri-foodstuffs importance. In this case, the system of the invention makes it possible to increase the copy number of this sequence per cell, and hence to increase the levels of production of this product and thus to increase the yields of the preparation process. The desired product can also be a peptide, polypeptide or protein participating in the biosynthesis (synthesis, degradation, transport or regulation) of a metabolite by the actinomycete strain in question. In this case, the system of the invention makes it possible to increase the copy number of this sequence per cell, and hence to increase the levels of production of this product, and thus either to increase the levels of production of the metabolite, or to block the biosynthesis of the metabolite, or to produce derivatives of the metabolite.

Plasmids comprising the site-specific integrating function of the invention can be used to permanently integrate copies of a heterologous gene of choice into the chromosome of many different hosts. The vectors can transform these hosts at a very high efficiency. Because the vectors do not have actinomycete origins of replication, the plasmids cannot exist as autonomously replicating vectors in actinomycete hosts. The plasmids only exist in their integrated form in these hosts. The integrated form is extremely stable which allows the gene copies to be maintained without antibiotic selective pressure. The result is highly beneficial in terms of cost, efficiency, and stability of the fermentation process.

Those skilled in the art will readily recognize that the variety of vectors which can be created that comprise this fragment is virtually limitless. The only absolute requirement is that the plasmid comprise an origin of replication which functions in the host cell in which

constructions are made, such as *E. coli* or *Bacillus*. No actinomycete origin of replication is required. In fact, in a specific embodiment the plasmid comprising the inetegrase comprises no actinomycete origin of replication. Other features, such as an antibiotic resistance gene, a multiple cloning site and cos site, are useful but not required. A description of the generation and uses of cosmid shuttle vectors can be found in Rao *et al.*, (Methods in Enzymology, 1987, 153:166-198). In short, any plasmid comprising the integrase is within the scope of this invention.

The integrating vectors can be used to integrate genes which increase the yield of known products or generate novel products, such as hybrid antibiotics or other novel secondary metabolites. The vector can also be used to integrate antibiotic resistance genes into strains in order to carry out bioconversions with compounds to which the strain is normally sensitive. The resulting transformed hosts and methods of making the antibiotics are within the scope of the present invention.

The integrase of the invention may thus be used in any actinomycete, in the genome of which the vector of the invention or its derivativesare is capable of integrating. In particular, they may be used in fermentation processes involving strains of Streptomyces, of mycobacteria, of bacilli, and the like. As an example, there may be mentioned the strains *S. pristinaespiralis* (ATCC 25486), *S.antibioticus* (DSM 40868), *S. bikiniensis* (ATCC 11062), *S. parvulus* (ATCC 12434), *S. glauescens* (ETH 22794), *S.actuosus* (ATCC 25421), *S. coelicolor* (A3(2)), *S. ambofaciens*, *S. lividans*, *S. griseofuscus*, *S. limosus*, and the like (*see also*, Smokvina *et al.*, Proceedings, 1:403-407).

In this connection, European Patent Publication No. EP 350,341 describes vectors derived from plasmid pSAM2 having very advantageous properties. These vectors are capable of integrating in a site-specific manner in the genome of actinomycetes, and possess a broad host range and high stability. Moreover, they may be used for transferring nucleic acids into actinomycetes and expressing these nucleic acids therein. U.S. Patent No. 5,741,675 describes tools capable of improving the conditions of industrial use of the vectors derived from pSAM2 by increasing the copy number of pSAM2 or its derivatives, since the free forms are present in a high copy number per cell. This patent also describes cassettes for the expression of this gene, vectors containing it and their use for inducing the appearance of free copies of pSAM2 or integrative vectors derived from the latter.

Alternatively, U.S. Patent No. 5,190,871 provides methods for increasing a given gene dosage and for adding heterologous genes that lead to the formation of new products such as

hybrid antibiotics using plasmids comprising the site-specific integrating function of phage phi.C31.

EXAMPLES

The following examples are provided for illustration purposes only and are not intended to limit the scope of the invention, which has been described in broad terms above.

EXAMPLE 1: Sequencing of Orsellinic Acid Synthetase

The DNA sequence of the *Micromonospora carbonaceae var. africana* (ATCC 39149) everninomicin biosynthetic region was obtained by sequencing inserts of recombinant DNA subclones containing contiguous or overlapping DNA segments of the region indicated in Figure 2A. All sequences representing the everninomicin region were fully contained in the overlapping cosmid clones pSPRX272, pSPRX262, pSPR192, pSPRX210, and pSPRX256 (Figure 2A). The sequence was obtained by subcloning and sequencing fragments bounded by restriction site as indicated in Figure 2A.

Preliminary sequences were also obtained for the cosmids pSPRX272 and pSPRX256. Restriction maps for these two cosmids are shown in Figure 2B and 2C, respectively. These restriction maps are characteristic of these two isolated cosmid clones of the *M. carbonaceae everninomicin* biosynthetic pathway or flanking regions thereof.

In order to obtain the *evrJ* gene, the sequence can be obtained by subcloning and sequencing of the fragments bounded by the *KpnI* sites at position 1, 25.9 kb, 29.6 kb, and 34.2 kb. The sequence can also be obtained by subcloning and sequencing of the fragments bounded by the *BamHI* sites at position 1, 24.5 kb, 27.0 kb, 28.8 kb and 30.5 kb. The resulting fragments should be ligated and cloned in an appropriate recombinant DNA vector. Clones containing the correct orientation of the fragment can be identified by restriction enzyme site mapping.

EXAMPLE 2: Transformation of *M. carbonacea* with pSPRH830

M. carbonacea was transformed with pSPRH830b (Figure 6) by conjugation from *E. coli* S17-1 (Mazodier *et al.*, Journal of Bacteriology, 1989, 6:3583-3585) to *M. carbonacea*. *E.coli* S17-1 containing pSPRH830b was grown overnight at 37 °C in LB supplemented with 100 μg/ml Ampicillin (Amp). The culture was inoculated into LB containing 100 μg/ml Amp at an 1:50 ratio and grown with shaking at 37 °C to an OD_{600} of 0.4 to 0.5. Cells were

harvested by centrifugation and washed three times with fresh LB lacking Amp. *M. carbonacea* was grown in TSB medium at 30 °C with shaking to stationary phase. *E. coli* S17-1 containing pSPRH830b prepared as described above was mixed with *M. carbonacea* in a total volume of 100 µl and plated on AS1 plates using a plastic hockey spreader. Plates were incubated for 15 hours at 29 °C and then overlaid with 50 µg/ml naladixic acid and 200 µg/ml Hygromycin for selection. Transconjugants appearing in 2-3 weeks were picked, homogenized and grown in TSB media with 50 µg/ml naladixic acid and 200 µg/ml hygromycin. Presence of pSPRH830b in *M. carbonacea* transformants was confirmed by PCR analysis and isolation of pSPRH830b from exconjugats.

The ability to transform *M. carbonacea* with pSPRH830b (on a multicopy plasmid) allows the introduction of second copies of genes contained in the everninomicin biosynthetic pathway or heterologous or mutated genes into *M. carbonacea*.

EXAMPLE 3: Transformation of *M. carbonacea* with pSPRH840

The pSPRH840 integrating vector (Figure 7A) was constructed as follows. A 4.0 kb *KpnII* fragment from the pSPR150 cosmid containing the *M.* carbonacea pMLP1 *intM* gene was ligated with *BamHI* cleaved pBluescriptII (Stratagene) to yield pSPRH819. Sequence analysis of the 4.0 kb *Kpn*I fragment from the cosmid revealed the presence of an integrase gene designated *intM*, an excisionase gene designated *xis*, and an integrase attachment site designated attP (Figure 7B).

BLAST analysis of intM showed homology to other integrases in the NRRL database. Analysis of the predicted attP site showed homology to the attP sites found phage phiC31 and plasmid pSAM2.

A 2.5 kb *NruI* to *XhoI* fragment from pSPR819 was treated with T4 polymerase to generate blunt DNA ends, alkaline phosphatase treated and ligated into the pCRTopo 2.1 vector (Invitrogen Corp, Carlsbad CA) to yield pSPRH853. A 2.6 kb *KpnI* to *PstI* fragment from pSPRH853 was ligated to *KpnI* and *PstI* digested pSPR826b (Figure 8) to yield pSPRH840 (Figure 7A). pSPRH840 was transformed into *M. carbonacea* SCC1413 and *M. halophitica* SCC760 as described in Example 2. Transconjugants appearing in two to three weeks were picked, homogenized, and grown in TSB medium supplimented with 50 µg/ml naladixic acid (Nacl) and 200 µg/ml Hygromycin. DNA was prepared from transconjugants, cleaved with *BamHI*, separated by gel electrophoresis, a Southern blot prepared, and probed with radiolabled pSPR826b. Southern hybridization analysis confirmed the presence of

pSPR826b sequences integrated into the *M. carbonacea* and *M. halophitica* chromosomes. Regions including pSPRH840 and chromosomal flanking sequences were cloned by digesting chromosomal DNA with *PstI* or *Kpn*I, ligating digested DNA and transforming *E. coli* XL10 (Stratagene, LaJolla, CA). *E. coli* transformants were isolated, plasmid DNA prepared and analyzed by digestion and gel electrophoresis. The attB/attP regions *M. carbonacea* and *M. halophitica* were each sequenced. Sequence analysis of this region confirmed that pSPRH840 had integrated into the *M. carbonacea* chromosome, specifically into a tRNA region (Figure 9A and 9B).

The ability to transform *M. carbonacea* with pSPRH840 allows the high frequency integration of second copies of genes contained in the everninomicin biosynthetic pathway or heterologous or mutated genes into *M. carbonacea*.

EXAMPLE 4: Overexpression and Isolation of Proteins From the EV Region

The coding region of *evrF* gene was amplified with PCR primers:

5' PR 657 CCC TCG AGA TGT CCA GCA AGA TCC TA (SEQ ID NO: 91);

3' PR 658 CGA ATT CTC AGG CAG ACT GCT CTG (SEQ ID NO: 92); and

5' PR 659: CCC TCG AGA ATG TCC AGC AAG ATC CTA (SEQ ID NO: 93);

3' PR 660: CGA ATT CAG ACT GCT CTG CCG CCG C (SEQ ID NO:94);

using the Advantage-GC Genomic PCR kit and Advantage HF polymerase (Clontech, Palo Alto, CA) and a Perkin-Elmer 9600 PCR machine (Foster City, CA). The 1.5 kb PCR products were digested with *XhoI* and *EcoRI* and the fragments were ligated to *XhoI* and *EcoRI* digested pBADHisA (primer pair PR657/PR658 product) and pBADMycHisC (primer pair PR659/PR660 product) and transformed into *E. coli* Top10 (Stratagene, LaJolla, CA). Transformants were analyzed by plasmid isolation followed by digestion and gel electrophoresis analysis. Appropriate clones were also verified by sequence analysis. This yielded the evrF expression clones pSPRE59 (pBADHisA) and pSPRE19 (pBADMycHisC). Top10 cells containing either pSPRE59 and pSPRE19 were grown overnight at 37 °C with shaking in LB containing 50ug/ml AMP. Overnight cultures were used to innoculate fresh LB containing 50 µg/ml and grown at 37 °C with shaking to an OD_{600} of 0.4 to 0.5. L-arabinose was added to a final concentration of 0.02% and the culture was incubated for an additional 4 hours. Cells were collected by centrifugation, resuspended in 100 µl Tris-Glycine buffer and boiled for five minutes. Whole cell protein lysate was loaded onto a SDS-PAGE gel, electrophoresed, and stained with coomassie blue to determine protein expression.

To isolate sufficient amounts of protein for raising antibodies, 100 ml of culture was processed as described above and the His-tagged EvrF protein was purified by Ni-NTA column chromatography using the Xpress Protein Purification System (Invitrogen, Carlsbad, CA). The recombinant EvrF protein was purified to over 90% homogeneity. This preparation was fractionated on SDS-PAGE gel, excised, and used to immunize New Zealand white rabbits to raise antibodies. Antisera were generated following standard protocol, *i.e.*, priming with complete Freund's adjuvant, (CFA) and boosting with incomplete Freund's adjuvant (IFA).

EXAMPLE 5: Everninomicin Pathway Expression of Putative Resistance Genes

Putative everninomicin resistance genes are expressed in the actinomycete vector pSPRH830b. Clones are obtained using standard molecular biology procedures. Plasmids are transformed into *Streptomyces lividans* or *Streptomyces griseofuscus* by PEG protoplast transformation or other standard actinomycete transformation procedures. Transformants are tested for increased resistance levels to everninomicin. A schematic of pSPRH830 the specific fragments to be cloned into is attached and shown in Figure 10.

The EV biosynthetic gene DNAs to be expressed by this recombinant vector are:

1) 4.9kb BamHI fragment containing
   - evrB, evrC - membrane pumps similar to mithramycin resistance.
2) 9.7kb HindIII/BamHI fragment containing
   - evbG, evbH - ABC transporter pumps, possible resistance mechanism.
3) 3.0kb BamHI fragment containing
   - evrE - MDR (Multiple drug resistance-type pump) transporter, possible resistance mechanism.
4) 3.56kb SacII fragment containing
   - evrY - dehalogenase, possible resistance mechanism
   - evrZ - muramidase/lysozyme homology, possible resistance mechansim.
5) 2.7kb BamHI fragment containing
   - evrMR - 23S rRNA methylase
6) A PCR fragment containing
   evcD and evcE - ABC transporters

Example 6: Insertional Inactivation of EV Pathway Genes

To confirm involvement of *evrJ*, (orsellinic acid synthetase) *evrF*, (halogenase) and

evrW (dTDP-glucose dehydratase) in EV biosynthesis these genes were disrupted in *M. carbonacea* via homologous recombination using the conjugative suicide vector pSPRH900b. Internal fragments of *evrJ*, *evrF*, and *evrW* were cloned into pSPRH900b to yield pSPRX572, pSPRX570, and pSPRX589 respectively. Plasmids pSPRX572, pSPRX570, and pSPRX589 were inserted into the chromosome by conjugation from *E. coli* into *M. carbonacea* to yield strains 572X, 570X and 589X repectively. Southern analysis confirmed insertion into the correct chromosomal loci for each plasmid. 572X, 570X and 589X strains showed a loss of EV production as shown by fermentation and analysis by HPLC indicating these genes are essential for EV production.

Production and determination of EV production was determined as follows. A mycelia stock of *M. carbonacea* was inoculated into the seed medium SIM-1 (10 ml) and incubated at 28 °C and 300 rpm. The seed inoculum (5ml) was then added to 4I + Co production medium (100 ml) and incubated at 28 °C and 300 rpm for 96 hours. A 10 ml aliquot of the fermentation broth was extracted with 20 ml of EtOAc, and the organic phase was evaporated to dryness. After resuspension in 2 ml of MeOH, 10 ml of the extract was subjected to HPLC analysis on a YMC-pack ODS-A C-18 column (3mm, 150 x 4.6 mm, Waters Corporation, Milford, MA). The column was equilibrated with 3mM tetramethyl ammonium hydroxide (pH to 7.2 with glacial acetic acid) with 70 % (vol/vol) MeOH and developed with a 24-min linear gradient from 70 to 90 % MeOH in the same 3mM tetramethyl ammonium hydroxide buffer at a flow rate of 0.8 ml/min. EV was detected at 270 nm by UV-Vis detection using a Agilent Series1100 HPLC system (Agilent Technologies).

* * * * *

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

It is further to be understood that all sizes and all molecular weight or molecular mass values are approximate, and are provided for description.

Patents, patent applications, procedures, and publications cited throughout this application are incorporated herein by reference in their entireties.